

# Gradient-Based Estimation of Uncertain Parameters for Elliptic Partial Differential Equations

Jeff Borggaard, Hans-Werner van Wyk

October 20, 2014

## Abstract

This paper addresses the estimation of uncertain distributed diffusion coefficients in elliptic systems based on noisy measurements of the model output. We formulate the parameter identification problem as an infinite dimensional constrained optimization problem for which we establish existence of minimizers as well as first order necessary conditions. A spectral approximation of the uncertain observations allows us to estimate the infinite dimensional problem by a smooth, albeit high dimensional, deterministic optimization problem, the so-called finite noise problem in the space of functions with bounded mixed derivatives. We prove convergence of finite noise minimizers to the appropriate infinite dimensional ones, and devise a stochastic augmented Lagrangian method for locating these numerically. Lastly, we illustrate our method with three numerical examples.

## 1 Introduction

This paper discusses a variational approach to estimating the parameter  $q$  in the elliptic system

$$-\nabla \cdot (q \nabla u) = f \text{ on } D, \quad u = 0 \text{ on } \partial D, \quad (1)$$

based on noisy measurements  $\hat{u}$  of  $u$ , when  $q$  is modeled as a spatially varying random field. Equation (1), defined over the physical domain  $D \subset \mathbb{R}^d$ , may describe the flow of fluid through a medium with permeability coefficient  $q$  or heat conduction across a material with conductivity  $q$ . Variational formulations in which the identification problem is posed as a constrained optimization, have been studied extensively for the case when  $q$  is deterministic [6, 12, 13, 19, 17]. Aleatoric uncertainty arising in these problems from imprecise, noisy measurements, variability in operating conditions, or unresolved scales are traditionally modeled as perturbations and addressed by means of regularization techniques. These approximate the original inverse problem by one in which the parameter depends continuously on the data  $\hat{u}$ , thus ensuring an estimation error commensurate with the noise level. However, when a statistical model for uncertainty in the dynamical system is available, it is desirable to incorporate this information more directly into the estimation framework to obtain an approximation not only of  $q$  itself but also of its probability distribution.

Bayesian methods provide a sampling-based approach to statistical parameter identification problems with random observations  $\hat{u}$ . By relating the observation noise in  $\hat{u}$  to the uncertainty associated with the estimated parameter via Bayes' Theorem [37, 38], these methods allow us to sample directly from the joint distribution of  $q$  at a given set of spatial points, through repeated

evaluation of the deterministic forward model. The convergence of numerical implementations of Bayesian methods, most notably Markov chain Monte Carlo schemes, depends predominantly on the statistical complexity of the input  $q$  and the measured output  $\hat{u}$  and is often difficult to assess. In addition, the computational cost of evaluating the forward model can possibly severely limit their efficiency.

There has also been a continued interest in adapting variational methods to estimate parameter uncertainty [5, 31, 32, 42]. Benefits include a well-established infrastructure of existing theory and algorithms, the possibility of incorporating multiple statistical influences, arising from uncertainty in boundary conditions or source terms for instance, and clearly defined convergence criteria. Let  $(\Omega, \mathcal{F}_{\hat{u}}, d\omega)$  be a complete probability space and suppose we have a statistical model of the measured data  $\hat{u}$  in the form of a random field  $\hat{u} = \hat{u}(x, \omega)$  contained in the tensor product  $\mathcal{H}_0^1(D) := H_0^1(D) \otimes L^2(\Omega)$ . A least squares formulation of the parameter identification problem in (1), when  $q(x, \omega)$  is a random field, may take the form

$$\begin{aligned} \min_{(q,u) \in \mathcal{H} \times \mathcal{H}_0^1} J(q, u) &:= \frac{1}{2} \|u - \hat{u}\|_{\mathcal{H}_0^1}^2 + \frac{\beta}{2} \|q\|_{\mathcal{H}}^2 \\ \text{s.t. } q &\in Q_{\text{ad}}, \quad e(q, u) = 0, \end{aligned} \quad (P)$$

where the regularization term with  $\beta > 0$  is added to ensure continuous dependence of the minimizer on the data  $\hat{u}$ . Here  $\mathcal{H} := H(D) \otimes L^2(\Omega)$ , where  $H(D)$  is any Hilbert space that imbeds continuously in  $L^\infty(D)$ , which may be taken to be the Sobolev space  $H^1(D)$  when  $d = 1$  or  $H^2(D)$  when  $d = 2, 3$  (see [19]). The feasible set  $Q_{\text{ad}}$  is given by

$$Q_{\text{ad}} = \{q \in \mathcal{H}(D) : 0 < q_{\min} \leq q(x, \omega) \text{ a.s. on } D \times \Omega, \|q(\cdot, \omega)\|_H \leq q_{\max} \text{ a.s. on } \Omega\},$$

while the stochastic equality constraint  $e(q, u) = 0$  represents Equation (1). It can also be written in its weak form as a functional equation  $\tilde{e}(q, u) = 0$  in  $\mathcal{H}^{-1}$ , where

$$\langle \tilde{e}(q, u), v \rangle_{\mathcal{H}^{-1}, \mathcal{H}_0^1} := \int_{\Omega} \int_D q(x, \omega) \nabla u(x, \omega) \cdot \nabla v(x, \omega) dx d\omega - \int_{\Omega} \int_D f(x) \phi(x, \omega) d\omega \quad (2)$$

for all  $v \in \mathcal{H}_0^1(D)$  [4]. For our purposes, it is useful to consider the equivalent functional equation  $e(q, u) = 0$  in  $\mathcal{H}_0^1$ , where  $e(q, u) := (-\Delta)^{-1} \tilde{e}(q, u)$  in the weak sense. Although these two forms of equality constraint are equivalent, pre-multiplication by the inverse Laplace operator adds a degree of preconditioning to the problem, as observed in [19]. We assume for the sake of simplicity that  $f \in L^2(D)$  is deterministic.

This formulation poses a number of theoretical, as well as computational challenges. The lack of smoothness of the random field  $q = q(x, \omega)$  in its stochastic component  $\omega$  limits the regularity of the equality constraint as a function of  $q$ , making it difficult to use theory analogous to the deterministic case in establishing first order necessary optimality conditions, as will be shown in Section 2. The most significant hurdle from a computational point of view is the need to approximate high dimensional integrals, both when evaluating the cost functional  $J$  and when dealing with the equality constraint (2). Monte Carlo type schemes seem inefficient, especially when compared with Bayesian methods. The recent success of Stochastic Galerkin methods [4, 41] and stochastic collocation-based approaches [4, 27] in efficiently estimating high dimensional integrals related to stochastic forward problems has, however, motivated investigations into their potential use in associated inverse and design problems.

In forward simulations, collocation methods make use of spectral expansions, such as the Karhunen-Loève (KL) series, to approximate the known input random field  $q$  by a smooth function of finitely many random variables, a so-called finite noise approximation. Standard PDE

regularity theory [4] then ensures that the corresponding model output  $u$  depends smoothly (even analytically) on these random variables. This facilitates the use of high-dimensional quadrature techniques, based on sparse grid interpolation of high order global polynomials. Inverse problems on the other hand are generally ill-posed and consequently any smoothness of a finite noise approximation of the given measured data  $\hat{u}$  does not necessarily carry over to the unknown parameter  $q$ . In variational formulations, explicit assumptions should therefore be made on the smoothness of finite noise approximations of  $q$  to facilitate efficient implementation, while also accurately estimating problem  $(P)$ .

We approximate  $(P)$  in the space of functions with bounded mixed derivatives. Posing the finite noise minimization problem  $(P^n)$  in this space not only guarantees that the equality constraint  $e(q, u)$  is twice Fréchet differentiable in  $q$  (see Section 4), but also allows for the use of numerical discretization schemes based on sparse grid hierarchical finite elements, approximations known not only for their amenability to adaptive refinement, but also for their effectiveness in mitigating the curse of dimensionality [11]. The authors in [42] demonstrate the use of piecewise linear hierarchical finite elements to approximate the finite noise design parameter in a least squares formulation of a heat flux control problem subject to system uncertainty, which is solved numerically through gradient-based methods. This paper aims to provide a rigorous framework within which to analyze and numerically approximate problems of the form  $(P)$ .

In Section 2, we establish existence and first order necessary optimality conditions for the infinite dimensional problem  $(P)$ . In Section 3 we make use of standard regularization theory to analytically justify the approximation of  $(P)$  by the finite noise problem  $(P^n)$ . We discuss existence and first order necessary optimality conditions for  $(P^n)$  in Section 4 and formulate an augmented Lagrangian algorithm for finding its solution in Section 5. Section 6 covers the numerical approximation of  $q$  and  $u$ , as well as the discretization of augmented Lagrangian optimization problem. Finally, we illustrate the application of our method on three numerical examples.

## 2 The Infinite Dimensional Problem

In order to accommodate the lack of smoothness of  $q$  as a function of  $\omega$  in our analysis, we impose inequality constraints uniformly in random space. Any function  $q$  in the feasible set  $Q_{\text{ad}}$ , satisfies the norm bound  $\|q(\cdot, \omega)\|_H \leq q_{\text{max}}$  uniformly on  $\Omega$ , which by the continuous imbedding of  $H(D)$  into  $L^\infty(D)$ , implies  $0 < q_{\text{min}} \leq q(x, \omega) \leq q_{\text{max}}$  for all  $(x, \omega) \in D \times \Omega$ . This assumption, while ruling out unbounded processes, nevertheless reflects actual physical constraints. The uniform coercivity condition  $0 < q_{\text{min}} \leq q(x, \omega)$ , guarantees that for each  $q \in Q_{\text{ad}}$ , there exists a unique solution  $u = u(q) \in \mathcal{H}_0^1(D)$  to the weak form (2) of the equality constraint  $e(q, u) = 0$  [3] satisfying the bound

$$\|u\|_{\mathcal{H}_0^1}^2 \leq \frac{C_D}{q_{\text{min}}} \|f\|_{L^2}. \quad (3)$$

Hence all  $q \in Q_{\text{ad}}$  and their respective model outputs  $u(q)$  have statistical moments of all orders.

### 2.1 Existence of Minimizers

An explicit stability estimate of  $u(q)$  in terms of the  $L^p(D \times \Omega)$  norm of  $q$  was given in [3, 4] for  $2 < p \leq \infty$ . These norms, besides not having Hilbert space structure, give rise to topologies

that are too weak for our purposes. The following lemmas establish the weak compactness of the feasible set, continuity of the solution mapping  $q \mapsto u(q)$  restricted to  $Q_{\text{ad}}$ , as well as the weak closedness of its graph in the stronger  $\mathcal{H}$  norm and will be used to prove the existence of solutions to (P).

**Lemma 2.1.** *The set  $Q_{\text{ad}}$  is closed, convex, and hence weakly compact in  $\mathcal{H}$ .*

*Proof.* Recall that

$$Q_{\text{ad}} = \{q \in \mathcal{H} : q(x, \omega) \geq q_{\min}, \text{ a.s. on } D \times \Omega, \|q(\cdot, \omega)\|_H \leq q_{\max} \text{ a.s. on } \Omega\}.$$

Convexity is easily verified. To show that  $Q_{\text{ad}}$  is closed, let  $\{q^n\} \subset Q_{\text{ad}}$  and  $q \in \mathcal{H}$  be such that

$$\|q^n - q\|_{\mathcal{H}}^2 = \int_{\Omega} \|q^n(\cdot, \omega) - q(\cdot, \omega)\|_H^2 d\omega \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since convergence in  $L^2(\Omega, d\omega)$  implies pointwise almost sure convergence of a subsequence on  $\Omega$ , it follows that

$$\|q^{n_k}(\cdot, \omega) - q(\cdot, \omega)\|_H \rightarrow 0 \quad \text{a.s. on } \Omega$$

for some subsequence  $\{q^{n_k}\} \subset Q_{\text{ad}}$ . Additionally,  $\|q^{n_k}(\cdot, \omega)\|_H \leq q_{\max}$  a.s. on  $\Omega$  for  $k \in \mathbb{N}$  and therefore  $q$  also satisfies this constraint. Finally,  $H(D)$  imbeds continuously in  $L^\infty(D)$ , which implies that the subsequence  $\{q^{n_k}\}$  in fact converges to  $q$  pointwise a.s. on  $D \times \Omega$ , ensuring that  $q$  also satisfies pointwise constraint  $q(x, \omega) \geq q_{\min}$  a.s. on  $D \times \Omega$ .  $\square$

**Lemma 2.2.** *The mapping  $u : q \in Q_{\text{ad}} \mapsto u(q) \in \mathcal{H}_0^1$  is continuous.*

*Proof.* Suppose  $q^n \rightarrow q$  in  $Q_{\text{ad}}$ . As in the proof of the previous lemma, there exists a subsequence  $q^{n_k} \rightarrow q$  pointwise a.s. on  $D \times \Omega$ . The upper bound on the function  $u$  established in [4, p. 1261] ensures that

$$\|u(q^{n_k}) - u(q)\|_{\mathcal{H}_0^1} \leq \left( \frac{C_D \|f\|_{L^2}}{q_{\min}^2} \right) \|q^{n_k} - q\|_{L^\infty(\Omega; L^\infty(D))} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where  $C_D$  is the constant appearing in the Poincaré inequality on  $D$ . Furthermore, since any subsequence of  $u(q^n)$  has a subsequence converging to  $u(q)$ , it follows that in fact  $u(q^n) \rightarrow u(q)$ .  $\square$

**Lemma 2.3.** *The graph  $\{(q, u) \in \mathcal{H} \times \mathcal{H}_0^1 : q \in Q_{\text{ad}}, u = u(q)\}$  of  $u$  is weakly closed.*

*Proof.* Let  $q^n$  be a sequence in  $Q_{\text{ad}}$ , so that  $q^n \rightharpoonup q$  in  $\mathcal{H}$  and  $u(q^n) \rightharpoonup u$  in  $\mathcal{H}_0^1$ . The weak compactness of  $Q_{\text{ad}}$  shown in Lemma 2.1, directly implies  $q \in Q_{\text{ad}}$ . It now remains to be shown that  $u = u(q)$  or equivalently that  $u$  solves  $e(q, u) = 0$ . Written in variational form, the requirement  $e(q, u) = 0$  is given by

$$\int_{\Omega} \int_D q \nabla u \cdot \nabla v \, dx \, d\omega = \int_{\Omega} \int_D f v \, dx \, d\omega \quad \text{for all } v \in \mathcal{H}_0^1. \quad (4)$$

Since the condition  $u^n = u(q^n) \Leftrightarrow e(q^n, u^n) = 0$  can be written as:

$$\int_{\Omega} \int_D q^n \nabla u^n \cdot \nabla v \, dx \, d\omega = \int_{\Omega} \int_D f v \, dx \, d\omega \quad \text{for all } v \in \mathcal{H}_0^1, \quad (5)$$

it suffices to show that the left hand side of (5) (or some subsequence thereof) converges to the left hand side of (4) for all  $v \in \mathcal{H}_0^1$ . Now for any  $n \geq 1$  and  $v \in \mathcal{H}_0^1$ ,

$$\begin{aligned} \int_{\Omega} \int_D (q^n \nabla u^n - q \nabla u) \cdot \nabla v \, dx \, d\omega &= \int_{\Omega} \int_D (q^n - q) \nabla u^n \cdot \nabla v \, dx \, d\omega \\ &\quad + \int_{\Omega} \int_D q \nabla (u^n - u) \cdot \nabla v \, dx \, d\omega. \end{aligned}$$

Let  $\{q^{n_k}\}$  be the subsequence of  $\{q^n\}$  that converges to  $q$  pointwise a.s. on  $D \times \Omega$ , as guaranteed by Lemma 2.1. We can then bound the first term by

$$\begin{aligned} &\left| \int_{\Omega} \int_D (q^{n_k} - q) \nabla u^{n_k} \cdot \nabla v \, dx \, d\omega \right| \\ &\leq \left( \int_{\Omega} \int_D |q^{n_k} - q| |\nabla u^{n_k}|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \left( \int_{\Omega} \int_D |q^{n_k} - q| |\nabla v|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \\ &\leq 2 \frac{q_{\max}}{q_{\min}} \|f\|_{L^2} \left( \int_{\Omega} \int_D |q^{n_k} - q| |\nabla v|^2 \, dx \, d\omega \right)^{\frac{1}{2}} \rightarrow 0 \quad \text{as } n_k \rightarrow \infty, \end{aligned}$$

by the Dominated Convergence Theorem, since the integrand is bounded above by  $2q_{\max} \|v\|_{\mathcal{H}_0^1}$ .

The second term in this sum converges to 0 due to the weak convergence  $u^n \rightharpoonup u$  and the fact that the mapping  $\|\cdot\|_q : u \mapsto \|u\|_q := \int_{\Omega} \int_D q |\nabla u|^2 \, dx \, d\omega$  defines a norm that is equivalent to  $\|\cdot\|_{\mathcal{H}_0^1}$ , by virtue of the fact that  $0 < q_{\min} \leq q(x, \omega) \leq q_{\max} < \infty$ . Therefore

$$\int_{\Omega} \int_D q \nabla u \cdot \nabla v \, dx \, d\omega = \lim_{n \rightarrow \infty} \int_{\Omega} \int_D q^n \nabla u^n \cdot \nabla v \, dx \, d\omega = \int_{\Omega} \int_D f v \, dx \, d\omega$$

for all  $v \in \mathcal{H}_0^1$  and hence  $e(q, u) = 0$ .  $\square$

By combining these lemmas, we can now show that a solution  $q^*$  of the infinite dimensional minimization problem (P) exists for any  $\beta \geq 0$ .

**Theorem 2.4** (Existence of Minimizers). *For each  $\beta \geq 0$ , the problem (P) has a minimizer.*

*Proof.* Let  $(q^n, u^n)$  be a minimizing sequence for the cost functional  $J$  over  $Q_{\text{ad}} \times \mathcal{H}_0^1$ , i.e.

$$\inf_{(q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1} J(q, u) = \lim_{n \rightarrow \infty} J(q^n, u^n) = \lim_{n \rightarrow \infty} \frac{1}{2} \|u^n - \hat{u}\|_{\mathcal{H}_0^1}^2 + \frac{\beta}{2} \|q^n\|_{\mathcal{H}}^2$$

Since  $u^n$  satisfies the equality constraint  $e(q^n, u^n) = 0$ , and consequently  $\|u^n\|_{\mathcal{H}_0^1} \leq \frac{1}{q_{\min}} \|f\|_{L^2}$  for all  $n \geq 1$  (Lax-Milgram), the Banach Alaoglu theorem guarantees the existence of a weakly convergent subsequence  $u^{n_k} \rightharpoonup u^* \in \mathcal{H}_0^1(D)$ . Moreover, the weak compactness of  $Q_{\text{ad}}$  established in Lemma 2.1 also yields a subsequence  $q^{n_k} \rightharpoonup q^*$  as  $k \rightarrow \infty$ , so that  $q^* \in Q_{\text{ad}}$ . The fact that the infimum of  $J$  is attained at the point  $(q^*, u^*)$  follows directly from the weak lower semicontinuity of norms [30]. Indeed,

$$\begin{aligned} J(q^*, u^*) &\leq \liminf_{n_1 \rightarrow \infty} \frac{1}{2} \|u^{n_1} - \hat{u}\|_{\mathcal{H}_0^1}^2 + \liminf_{n_1 \rightarrow \infty} \frac{\beta}{2} \|q^{n_1}\|_{\mathcal{H}}^2 \\ &\leq \liminf_{n_1 \rightarrow \infty} \left( \frac{1}{2} \|u^{n_1} - \hat{u}\|_{\mathcal{H}_0^1}^2 + \frac{\beta}{2} \|q^{n_1}\|_{\mathcal{H}}^2 \right) = \inf_{(q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1} J(q, u). \end{aligned}$$

Finally, it follows directly from Lemma 2.3 that  $u^* = u^*(q^*)$  and hence  $u^*$  satisfies the inequality constraint  $e(q^*, u^*) = 0$ . The regularization term was not required to show the existence of minimizers.  $\square$

## 2.2 A Saddle Point Condition

Although solutions to  $(P)$  exist, the inherent lack of smoothness of  $q$  in the stochastic variable  $\omega$  complicates the establishment of traditional necessary optimality conditions. A short calculation reveals that the equality constraint  $e(q, u) = 0$  is not Fréchet differentiable, as a function  $q$  in  $\mathcal{H}$ . Additionally, the set of constraints has an empty interior in the  $\mathcal{H}$ -norm. Instead, we follow [14] in deriving a saddle point condition for the optimizer  $(q^*, u^*)$  of  $(P)$  with the help of a Hahn-Banach separation argument.

Let  $\langle \cdot, \cdot \rangle$  denote the  $L^2(D \times \Omega)$  inner product. For any triple  $(q, u, \lambda) \in \mathcal{H} \times \mathcal{H}_0^1 \times \mathcal{H}_0^1$ , we define the Lagrangian functional by

$$L(q, u, \lambda) = J(q, u) + \langle e(q, u), \lambda \rangle_{\mathcal{H}_0^1} = \frac{1}{2} \|u - \hat{u}\|_{\mathcal{H}_0^1}^2 + \frac{\beta}{2} \|q\|_{\mathcal{H}}^2 + \langle q \nabla u, \nabla \lambda \rangle - \langle f, \lambda \rangle.$$

The main theorem of this subsection is the following

**Theorem 2.5** (Saddle Point Condition). *Let  $(q^*, u^*) \in Q_{\text{ad}} \times \mathcal{H}_0^1$  solve problem  $(P)$ . Then there exists a Lagrange multiplier  $\lambda^* \in \mathcal{H}_0^1$  so that the saddle point condition*

$$L(q^*, u^*, \mu) \leq L(q^*, u^*, \lambda^*) \leq L(q, u, \lambda^*) \quad (6)$$

holds for all  $(q, u, \mu) \in Q_{\text{ad}} \times \mathcal{H}_0^1 \times \mathcal{H}_0^1$ .

*Proof.* Note that the second inequality simply reflects the optimality of  $(q^*, u^*)$ . To obtain the first inequality, we rely on a Hahn-Banach separation argument. Let

$$S = \{(J(q, u) - J(q^*, u^*) + s, e(q, u)) \in \mathbb{R} \times \mathcal{H}_0^1 : (q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1, s \geq 0\}$$

and

$$T = \{(-t, 0) \in \mathbb{R} \times \mathcal{H}_0^1 : t > 0\}$$

In the ensuing three lemmas we will show that

1.  $S$  and  $T$  are convex (Lemma 2.6),
2.  $S \cap T = \emptyset$  (Lemma 2.7), and
3.  $S$  has at least one interior point (Lemma 2.8).

The Hahn-Banach Theorem thus gives rise to a separating hyperplane, i.e. a pair  $(\alpha_0, \lambda_0) \neq (0, 0)$  in  $\mathbb{R} \times \mathcal{H}_0^1$ , such that

$$\alpha_0(J(q, u) - J(q^*, u^*) + s) + \langle e(q, u), \lambda_0 \rangle_{\mathcal{H}_0^1} \geq -t\alpha_0 \quad \forall t > 0, s \geq 0, (q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1. \quad (7)$$

Letting  $s = t = 1$  and  $(q, u) = (q^*, u^*)$  readily yields  $\alpha_0 \geq 0$ . In fact  $\alpha_0 > 0$ . Suppose to the contrary that  $\alpha_0 = 0$ . Then by (7)

$$\langle e(q, u), \lambda_0 \rangle_{\mathcal{H}_0^1} = \langle q \nabla u, \nabla \lambda_0 \rangle - \langle f, \lambda_0 \rangle \geq 0 \quad \forall (q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1$$

particularly for  $q = q^*$  and  $u \in \mathcal{H}_0^1$  satisfying  $\langle q^* \nabla u, \nabla \phi \rangle - \langle f - \lambda_0, \phi \rangle = 0 \quad \forall \phi \in \mathcal{H}_0^1$ , we have

$$\langle q^* \nabla u, \nabla \lambda_0 \rangle - \langle f, \lambda_0 \rangle = \langle f - \lambda_0, \lambda_0 \rangle - \langle f, \lambda_0 \rangle = -\langle \lambda_0, \lambda_0 \rangle \geq 0,$$

which implies that  $\lambda_0 = 0$ . This contradicts the fact that  $(\alpha_0, \lambda_0) \neq (0, 0)$ . Dividing (7) by  $\alpha_0$  and letting  $\lambda^* = \lambda_0/\alpha_0$  yields  $J(q^*, u^*) \leq J(q, u) + \langle e(q, u), \lambda^* \rangle_{\mathcal{H}_0^1} \quad \forall (q, u) \in Q_{\text{ad}} \times \mathcal{H}_0^1$  and hence

$$\begin{aligned} L(q^*, u^*, \mu) &= J(q^*, u^*) + \langle e(q^*, u^*), \mu \rangle_{\mathcal{H}_0^1} = J(q^*, u^*) \\ &\leq J(q, u) + \langle e(q, u), \lambda^* \rangle_{\mathcal{H}_0^1} = L(q, u, \lambda^*) \end{aligned}$$

for all  $(q, u, \mu) \in Q_{\text{ad}} \times \mathcal{H}_0^1 \times \mathcal{H}_0^1$ . □

**Lemma 2.6.** *The sets  $S$  and  $T$  are convex.*

*Proof.* Clearly,  $T$  is convex. Let  $0 \leq \alpha \leq 1$  and consider the convex combination  $P_\alpha = \alpha P_1 + (1 - \alpha)P_2$  where  $P_1, P_2 \in S$ . Hence  $P_\alpha$  is of the form  $P_\alpha = (p_\alpha, w_\alpha)$  where

$$\begin{aligned} p_\alpha &= \alpha(J(q_1, u_1) - J(q^*, u^*) + s_1) + (1 - \alpha)(J(q_2, u_2) - J(q^*, u^*) + s_2) \\ w_\alpha &= \alpha e(q_1, u_1) + (1 - \alpha)e(q_2, u_2) \end{aligned}$$

with  $q_1, q_2 \in Q_{\text{ad}}$ ,  $u_1, u_2 \in \mathcal{H}_0^1$ , and  $s_1, s_2 \geq 0$ . It now remains to show that  $w_\alpha = e(q_\alpha, u_\alpha)$  for some  $(q_\alpha, u_\alpha) \in Q_{\text{ad}} \times \mathcal{H}_0^1$  and  $p_\alpha = J(q_\alpha, u_\alpha) - J(q^*, u^*) + s_\alpha$  for some  $s_\alpha \geq 0$ . Let  $q_\alpha = \alpha q_1 + (1 - \alpha)q_2 \in Q_{\text{ad}}$  and let  $u_\alpha \in \mathcal{H}_0^1$  be the unique solution of the variational problem

$$\langle q_\alpha \nabla u_\alpha, \nabla \phi \rangle = \langle \alpha q_1 \nabla u_1 + (1 - \alpha)q_2 \nabla u_2, \nabla \phi \rangle \quad \forall \phi \in \mathcal{H}_0^1.$$

Therefore

$$\begin{aligned} \langle w_\alpha, \phi \rangle_{\mathcal{H}_0^1} &= \langle \alpha q_1 \nabla u_1 + (1 - \alpha)q_2 \nabla u_2, \nabla \phi \rangle - \langle f, \phi \rangle \\ &= \langle q_\alpha \nabla u_\alpha, \nabla \phi \rangle - \langle f, \phi \rangle = \langle e(q_\alpha, u_\alpha), \phi \rangle_{\mathcal{H}_0^1} \quad \forall \phi \in \mathcal{H}_0^1 \end{aligned}$$

which implies that  $w_\alpha = e(q_\alpha, u_\alpha)$ . Moreover, it follows readily from the convexity of norms that

$$J(q_\alpha, u_\alpha) \leq \alpha J(q_1, u_1) + (1 - \alpha)J(q_2, u_2)$$

and therefore letting

$$s_\alpha = \alpha J(q_1, u_1) + (1 - \alpha)J(q_2, u_2) - J(q_\alpha, u_\alpha) + \alpha s_1 + (1 - \alpha)s_2 \geq \alpha s_1 + (1 - \alpha)s_2 \geq 0$$

we obtain

$$p_\alpha = J(q_\alpha, u_\alpha) - J(q^*, u^*) + s_\alpha.$$

□

**Lemma 2.7.** *The sets  $S$  and  $T$  are disjoint.*

*Proof.* This follows directly from the fact that  $J(q, u) \geq J(q^*, u^*)$  for all points  $(q, u)$  in  $Q_{\text{ad}} \times \mathcal{H}_0^1$  □

**Lemma 2.8.** *The set  $S$  has a non-empty interior.*

*Proof.* Clearly  $(s_0, 0) = (J(q^*, u^*) - J(q^*, u^*) + s_0, e(q^*, u^*)) \in S$  for any  $s_0 > 0$ . For any  $\epsilon \in (0, 1)$ , let  $(s, w)$  belong to the  $\epsilon$ -neighborhood of  $(s_0, 0)$ . In other words  $|s - s_0| + \|w\|_{\mathcal{H}_0^1} \leq \epsilon$ . Let  $q = q^*$  and let  $u$  be the solution to the problem

$$\langle q^* \nabla u, \nabla \phi \rangle = \langle f, \phi \rangle + \langle \nabla w, \nabla \phi \rangle \quad \forall \phi \in \mathcal{H}_0^1(D). \quad (8)$$

Clearly,  $w = e(q^*, u)$  by definition. Then

$$\begin{aligned} s' &:= s_0 + J(q^*, u^*) - J(q, u) = s_0 + J(q^*, u^*) - J(q^*, u) \\ &= s_0 + \frac{1}{2} \int_\Omega \int_D |\nabla(u^*(x, \omega) - \hat{u}(x, \omega))|^2 dx d\omega - \frac{1}{2} \int_\Omega \int_D |\nabla(u(x, \omega) - \hat{u}(x, \omega))|^2 dx d\omega \\ &= s_0 - \frac{1}{2} \int_\Omega \int_D \nabla(u(x, \omega) - u^*(x, \omega)) \cdot \nabla(u(x, \omega) + u^*(x, \omega) - 2\hat{u}(x, \omega)) dx d\omega \end{aligned}$$

Now  $u^*$  satisfies  $e(q^*, u^*) = 0$  and hence  $\|u^*\|_{\mathcal{H}_0^1} \leq \frac{C_D}{q_{\min}} \|f\|_{L^2}$  by (3). Similarly, since  $u$  solves (8), it follows that  $\|u\|_{\mathcal{H}_0^1} \leq \frac{C_D}{q_{\min}} (\|f\|_{L^2} + \|w\|_{\mathcal{H}_0^1}) \leq \frac{C_D}{q_{\min}} (\|f\|_{L^2} + \epsilon)$  and hence  $\|u - u^*\|_{\mathcal{H}_0^1} \leq \frac{C_D}{q_{\min}} \epsilon$ . We therefore have

$$\begin{aligned} s' &\geq s_0 - \frac{1}{2} \|u - u^*\|_{\mathcal{H}_0^1} (\|u^*\|_{\mathcal{H}_0^1} + \|u\|_{\mathcal{H}_0^1} + 2\|\hat{u}\|_{\mathcal{H}_0^1}) \\ &\geq s_0 - \frac{\epsilon}{2q_{\min}} \left( \frac{C_D}{q_{\min}} \|f\|_{L^2} + \frac{C_D}{q_{\min}} (\|f\|_{L^2} + \epsilon) + 2\|\hat{u}\|_{\mathcal{H}_0^1} \right) \\ &\geq s_0 - \frac{\epsilon}{2q_{\min}^2} (2C_D \|f\|_{L^2} + C_D \epsilon + 2q_{\min} \|\hat{u}\|_{\mathcal{H}_0^1}) \geq 0 \end{aligned}$$

for small enough  $\epsilon > 0$ . Therefore  $(s, w) = (J(q^*, u) - J(q^*, u^*) + s', e(q^*, u)) \in S$  for any  $(s, w)$  in a small enough  $\epsilon$ -neighborhood of  $(s_0, 0)$ .  $\square$

In the following section, we will show that if the observed data  $\hat{u}$  is expressed as a Karhunen-Loève series [23, 33], we may approximate problem  $(P)$  by a finite noise optimization problem  $(P^n)$ , where  $q$  is a smooth, albeit high-dimensional, function of  $x$  and intermediary random variables  $\{Y_i\}_{i=1}^n$ . The convergence framework not only informs the choice of numerical discretization, but also suggests the use of a dimension-adaptive scheme to exploit the progressive ‘smoothing’ of the problem.

### 3 Approximation by the Finite Noise Problem

According to [23], the random field  $\hat{u}$  may be written as the Karhunen-Loève (KL) series

$$\hat{u}(x, \omega) = \hat{u}_0(x) + \sum_{k=1}^{\infty} \sqrt{\nu_k} b_k(x) Y_k(\omega), \quad (9)$$

where  $\{Y_k\}_{k=1}^{\infty}$  is an uncorrelated orthonormal sequence of random variables with zero mean and unit variance and  $(\nu_k, b_k)$  is the eigenpair sequence of  $\hat{u}$ ’s compact covariance operator  $\mathcal{C}_{\hat{u}} : H_0^1(D) \rightarrow H_0^1(D)$  [33]. Moreover, the truncated series

$$\hat{u}^n(x, \omega) = u_0(x) + \sum_{k=1}^n \sqrt{\nu_k} b_k(x) Y_k(\omega)$$

converges to  $\hat{u}$  in  $\mathcal{H}_0^1$ , i.e.  $\|\hat{u} - \hat{u}^n\|_{\mathcal{H}_0^1} \rightarrow 0$  as  $n \rightarrow \infty$ . Assume w.l.o.g. that  $\{Y_i\}_{i=1}^{\infty}$  forms a complete orthonormal basis for  $L_0^2(\Omega)$ , the set of functions in  $L^2(\Omega)$  with zero mean. If this is not the case, we can restrict ourselves to  $L_0^2(\Omega) \cap \text{span}\{Y_i\}$ . The following additional assumption imposes restrictions on the range of the random vectors we consider.

**Assumption 3.1.** Assume the random variables  $\{Y_n\}$  are bounded uniformly in  $n$ , i.e.

$$y_{\min} \leq Y_n(\omega) \leq y_{\max} \quad \text{a.s. on } \Omega \text{ for all } n \in \mathbb{N} \text{ and some } y_{\min}, y_{\max} \in \mathbb{R}.$$

Furthermore, assume that for any  $n$  the probability measure of the random vector  $Y = (Y_1, \dots, Y_n)$  is absolutely continuous with respect to the Lebesgue measure and hence  $Y$  has joint density  $\rho_n : \Gamma^n \rightarrow [0, \infty)$ , where the hypercube  $\Gamma^n := \prod_{i=1}^n \Gamma_i \subset [y_{\min}, y_{\max}]^n$  denotes the range of  $Y$ .

Since  $\hat{u}^n$  depends on  $\omega$  only through the intermediary variables  $\{Y_i\}_{i=1}^n$ , it seems reasonable to also estimate the unknown parameter  $q^n$  as a function of these, i.e.

$$q^n(x, \omega) = q^n(x, Y_1(\omega), \dots, Y_n(\omega)).$$



The appropriate parameter space for the finite noise identification problem is not immediately apparent. In order for the finite noise optimization problem to approximate  $(P)$ ,  $q_n$  should at the very least be square integrable in  $y$ , i.e.  $q^n \in \tilde{H}(D) := H(D) \otimes L^2(\Gamma^n) \subset \mathcal{H}(D)$ . With this parameter space, however, the finite noise problem suffers from the same lack of regularity encountered in the infinite dimensional problem  $(P)$ . In order to ensure both that the finite noise equality constraint  $e_n(q, u) = 0$  is Fréchet differentiable and that the set of admissible parameters  $Q_{\text{ad}}$  has a non-empty interior, we require a higher degree of smoothness in  $q$  as a function of  $y \in \Gamma^n$ .

For the sake of our analysis, we therefore seek finite noise minimizers  $q_n^*$  in the space  $\tilde{H}_{\text{mix}} := H(D) \otimes H_{\text{mix}}^s(\Gamma^n)$ , where  $H_{\text{mix}}^s(\Gamma^n)$  is the space of functions with bounded mixed derivatives,  $s \geq 1$  [39]. A function  $v \in \tilde{H}_{\text{mix}} \subset L^2(D \times \Gamma^n)$  is one for which the  $\tilde{H}_{\text{mix}}$ -norm,

$$\|v\|_{\tilde{H}_{\text{mix}}}^2 := \sum_{|\gamma|_\infty \leq s} \sum_{|\alpha|_1 \leq t_d} \int_D \int_{\Gamma^n} |D_y^\gamma D_x^\alpha v(x, y)|^2 \rho_n(y) dy dx \quad (10)$$

is finite, where  $\gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{N}^n$  and  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$  are multi-indices, with  $|\gamma|_\infty = \max\{\gamma_1, \dots, \gamma_n\}$ ,  $|\alpha|_1 = \alpha_1 + \dots + \alpha_d$  and  $t_d = 1$  when  $d = 1$  or  $t_d = 2$  when  $d = 2, 3$ . Apart from considerations of convenience, the use of this parameter space is partly justified by the fact that  $\{Y_n\}_{n=1}^\infty$  forms a basis for  $L_0^2(\Omega)$ . The minimizer  $q^*$  of the original infinite dimensional problem  $(P)$  thus takes the form

$$q^*(x, \omega) = q_0^*(x) + \sum_{n=1}^\infty q_n(x) Y_n(\omega),$$

which is linear in each of the random variables  $Y_n$ . Any minimizer  $q_n^*$  of  $(P^n)$  that approximates  $q^*$  (even in the weak sense) is therefore expected to depend relatively smoothly on  $y$  when  $n$  is large. At low orders of approximation, on the other hand, the parameter  $q$  that gives rise to the model output  $u(q)$  most closely resembling the partial data  $\hat{u}^n$  may not exhibit the same degree of smoothness in the variable  $y = (y_1, \dots, y_n)$ . Since the accuracy in approximation of functions in high dimensions benefits greatly from a high degree of smoothness [7], this suggests the use of a dimension adaptive strategy in which the smoothness requirement of the parameter is gradually strengthened as the stochastic dimension  $n$  increases.

We can now proceed to formulate a finite noise least squares parameter estimation problem for the perturbed, finite noise data  $\hat{u}^n$ :

$$\begin{aligned} \min_{(q, u) \in \tilde{H}_{\text{mix}} \times \tilde{H}_0^1} J(q, u) &:= \frac{1}{2} \|u - \hat{u}^n\|_{\tilde{H}_0^1}^2 + \frac{\beta_n}{2} \|q\|_{\tilde{H}_{\text{mix}}}^2 \\ \text{s.t. } q &\in Q_{\text{ad}}^n, \quad e_n(q, u) = 0 \end{aligned} \quad (P^n)$$

where  $e_n(\cdot, \cdot) : \tilde{H}_{\text{mix}} \times \tilde{H}_0^1 \rightarrow \tilde{H}_0^1$  is defined by  $e_n(q, u) = (-\Delta)^{-1} \tilde{e}_n(q, u)$  with

$$\begin{aligned} \langle \tilde{e}_n(q, u), v \rangle_{\tilde{H}^{-1}, \tilde{H}_0^1} &:= \int_{\Gamma^n} \int_D q(x, y) \nabla u(x, y) \cdot \nabla v(x, y) \rho_n(y) dx dy \\ &\quad - \int_{\Gamma^n} f(x, y) v(x, y) \rho_n(y) dx dy \end{aligned}$$

for all  $v \in \tilde{H}_0^1(D)$ , and

$$Q_{\text{ad}}^n := \left\{ q \in \tilde{H}^n : \begin{array}{ll} 0 < q_{\min} - \frac{1}{k_n} \leq q(x, y) & \text{a.s. on } D \times \Gamma^n, \\ \|q(\cdot, y)\|_H \leq q_{\max} + \frac{1}{k_n} & \text{a.s. on } \Gamma^n \end{array} \right\}.$$

with  $k_n \rightarrow \infty$  a monotone increasing approximation parameter to be specified later.

In the following, we justify the use of this approximation scheme by demonstrating that it not only lends itself more readily to standard first- and second-order optimization theory, but also that  $(P^n)$  approximates  $(P)$  in a certain sense. In particular, we first show that, as  $n \rightarrow \infty$  and  $\beta_n \rightarrow 0$ , the sequence of minimizers  $q_n^*$  of problem  $(P^n)$  has a weakly convergent subsequence and that the limits of all convergent subsequences minimize the infinite dimensional problem  $(P)$ . Tikhonov regularization theory for non-linear least squares problems [8] provides the theoretical framework underlying the arguments in this section.

In order to mediate between the minimizer  $q_n^*$  of the finite noise problem  $(P^n)$ , formulated in the  $\tilde{H}_{\text{mix}}$  norm, and that of the infinite dimensional problem, whose minimizer  $q^*$  is measured in the  $\mathcal{H}$  norm, we make use of the projection of  $q^*$  on the first  $n$  basis vectors:

$$\mathbf{P}^n q^* = q_0^*(x) + \sum_{i=1}^n q_i(x) Y_i(\omega).$$

Evidently,  $\mathbf{P}^n q^* \rightarrow q^*$  as  $n \rightarrow \infty$  in  $\mathcal{H}$ . Moreover, seeing that  $\mathbf{P}^n q^*$  is linear in  $y$ , it's norm in  $\tilde{H}_{\text{mix}}$  can be bounded in terms of its norm in  $\mathcal{H}$  as the following computation shows:

**Lemma 3.2.**

$$\|\mathbf{P}^n q^*\|_{\tilde{H}_{\text{mix}}} \leq \sqrt{2} \|\mathbf{P}^n q^*\|_{\mathcal{H}}.$$

*Proof.* Let  $e_i$  be the  $i^{\text{th}}$  standard basis vector for  $\mathbb{N}^n$ . We now apply expression (10) to  $\mathbf{P}^n q^*$  to obtain

$$\begin{aligned} \|\mathbf{P}^n q^*\|_{\tilde{H}_{\text{mix}}}^2 &:= \sum_{|\gamma|_\infty \leq s} \sum_{|\alpha|_1 \leq t_d} \int_D \int_{\Gamma^n} \left| D_y^\gamma D_x^\alpha \left[ q_0(x) + \sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\ &= \sum_{|\alpha|_1 \leq t_d} \int_D \int_{\Gamma^n} \left| D_y^0 D_x^\alpha \left[ q_0(x) + \sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\ &\quad + \sum_{i=1}^n \sum_{|\alpha|_1 \leq t_d} \int_D \int_{\Gamma^n} \left| D_y^{e_i} D_x^\alpha \left[ \sum_{i=1}^n q_i(x) y_i \right] \right|^2 \rho_n(y) dy dx \\ &= \int_{\Gamma^n} \|\mathbf{P}^n q^*(\cdot, \omega)\|_H^2 \rho_n(y) dy + \sum_{i=1}^n \sum_{|\alpha|_1 \leq t_d} \int_D \int_{\Gamma^n} |D_x^\alpha q_i(x)|^2 \rho_n(y) dy dx \\ &= \|\mathbf{P}^n q^*\|_H^2 + \sum_{i=1}^n \|q_i\|_H^2 = 2 \sum_{i=0}^n \|q_i\|_H^2 - \|q_0\|_H^2 \leq 2 \|\mathbf{P}^n q^*\|_{\mathcal{H}}^2. \end{aligned}$$

The second and third equalities follow from the fact that

$$D_y^\gamma \left[ \sum_{i=1}^n q_i(x) y_i \right] = \begin{cases} \sum_{i=1}^n q_i(x) y_i & , \text{ if } \gamma = 0 \\ q_i(x) & , \text{ if } \gamma = e_i \\ 0 & , \text{ otherwise} \end{cases}.$$

□

The next lemma addresses the feasibility of  $\mathbf{P}^n q^*$ . Although  $\mathbf{P}^n q^*$  does not necessarily lie in the feasible region  $Q_{\text{ad}}$ , the set on which  $\mathbf{P}^n q^* \notin Q_{\text{ad}}$  can be made arbitrarily small as  $n \rightarrow \infty$ . Let  $\mathcal{A}_n$  be the event that  $\mathbf{P}^n q^*$  lies inside the approximate feasible region  $Q_{\text{ad}}^n$ , i.e.

$$\mathcal{A}_n := \left\{ \omega \in \Omega : 0 < q_{\min} - \frac{1}{k_n} \leq \mathbf{P}^n q^*(x, \omega) \text{ a.s. on } D, \|\mathbf{P}^n q^*(\cdot, \omega)\|_H \leq q_{\max} + \frac{1}{k_n} \right\}.$$

Then we have

**Lemma 3.3.** *There is a monotonically increasing sequence  $k_n \rightarrow \infty$  so that  $\mathbb{P}(\Omega \setminus \mathcal{A}_n) \leq \frac{1}{k_n}$  for all  $n \in \mathbb{N}$ .*

*Proof.* For any  $n \geq 1$ , let  $k_n$  satisfy  $\|\mathbf{P}^n q^* - q^*\|_{\mathcal{H}}^2 = \frac{1}{C^2 k_n^3}$ , where  $C \geq 1$  is the imbedding constant for  $H(D) \hookrightarrow L^\infty(D)$ . Clearly  $k_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let

$$\mathcal{B}_n = \left\{ \omega \in \Omega : \|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H \leq \frac{1}{C k_n} \right\}.$$

For any  $\omega \in \mathcal{B}_n$ ,

$$\left| \|\mathbf{P}^n q^*(\cdot, \omega)\|_H - \|q^*(\cdot, \omega)\|_H \right| \leq \|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H \leq \frac{1}{C k_n} \leq \frac{1}{k_n}$$

and

$$\|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_{L^\infty} \leq C \|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H \leq \frac{1}{k_n},$$

which implies  $\mathcal{B}_n \subset \mathcal{A}_n$ . Moreover, according to Chebychev's inequality

$$\mathbb{P}(\Omega \setminus \mathcal{A}_n) \leq \mathbb{P}(\Omega \setminus \mathcal{B}_n) \leq C^2 k_n^2 \int_{\Omega} \|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H^2 d\omega = C^2 k_n^2 \|\mathbf{P}^n q^* - q^*\|_{\mathcal{H}}^2 \leq \frac{1}{k_n}.$$

□

In order to ensure strict adherence to the inequality constraints of  $(P^n)$  for every  $n$ , we modify  $\mathbf{P}^n q^*(\cdot, \omega)$  on  $\Omega \setminus \mathcal{A}_n$ .

**Definition 3.4.** For all  $n \in \mathbb{N}$ , let  $\hat{q}_n^* \in \tilde{H}_{\text{mix}} \subset \mathcal{H}$  be defined as follows:

$$\hat{q}_n^* := \begin{cases} \mathbf{P}^n q^*, & \omega \in \mathcal{A}_n \\ q_n^*, & \omega \notin \mathcal{A}_n \end{cases}. \quad (11)$$

Evidently  $\hat{q}_n^* \in Q_{\text{ad}} \cap \tilde{H}_{\text{mix}}$  and in light of Lemma 3.3, it is reasonable to expect  $\hat{q}_n^* \approx \mathbf{P}^n q^*$  for large  $n$ , except on sets of negligible measure. Indeed

**Lemma 3.5.**  $\hat{q}_n^* \rightarrow q^*$  in  $\mathcal{H}$  as  $n \rightarrow \infty$ .

*Proof.*

$$\begin{aligned} \|\hat{q}_n^* - q^*\|_{\mathcal{H}} &= \int_{\mathcal{A}_n} \|\mathbf{P}^n q^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H^2 d\omega + \int_{\Omega \setminus \mathcal{A}_n} \|q_n^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H^2 d\omega \\ &\leq \|\mathbf{P}^n q^* - q^*\|_{\mathcal{H}}^2 + \mathbb{P}(\Omega \setminus \mathcal{A}_n) \sup_{\omega \in \Omega} \|q_n^*(\cdot, \omega) - q^*(\cdot, \omega)\|_H^2 \\ &\leq \|\mathbf{P}^n q^* - q^*\|_{\mathcal{H}}^2 + \frac{1}{k_n} 4(q_{\max} + \frac{1}{k_1})^2 \rightarrow 0. \end{aligned}$$

□

We are now in a position to prove the main theorem of this section. For its proof we will make use of the fact that, due to the lower semicontinuity of norms

$$x_n \rightharpoonup x, \quad \limsup_{n \rightarrow \infty} \|x_n\| \leq \|x\| \Rightarrow x_n \rightarrow x \quad (12)$$

for any sequence  $x_n$  in a Hilbert space.

**Theorem 3.6.** *Let  $\|\hat{u} - \hat{u}^n\|_{\mathcal{H}_0^1} \rightarrow 0$  and  $\beta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then the sequence of minimizers  $q_n^*$  of  $(P^n)$  has a subsequence converging weakly to a minimizer of infinite dimensional problem  $(P)$  and the limit of every weakly convergent subsequence is a minimizer of  $(P)$ . The corresponding model outputs converge strongly to the infinite dimensional minimizer's model output.*

*Proof.* Since  $q_n^*$  is optimal for  $(P^n)$ , we have

$$\|u(q_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 + \beta_n \|q_n^*\|_{\tilde{H}_{\text{mix}}}^2 \leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 + \beta_n \|\hat{q}_n^*\|_{\tilde{H}_{\text{mix}}}^2. \quad (13)$$

Moreover, by definition  $\hat{q}_n^*(\cdot, Y(\omega)) = q_n^*(\cdot, Y(\omega))$  for all  $Y \in Y(\Omega \setminus \mathcal{A}_n)$  and hence

$$\begin{aligned} & \|\hat{q}_n^*\|_{\mathcal{H}}^2 - \|q_n^*\|_{\mathcal{H}}^2 \\ &= \sum_{|\gamma|_\infty \leq 1} \sum_{|\alpha|_1 \leq t_d} \left( \int_{Y(\mathcal{A}_n)} \int_D |D_y^\gamma D_x^\alpha \mathbf{P}^n q^*|^2 \rho_n(y) dx dy - \int_{Y(\mathcal{A}_n)} \int_D |D_y^\gamma D_x^\alpha q_n^*|^2 \rho_n(y) dx dy \right) \\ &\leq \sum_{|\gamma|_\infty \leq 1} \sum_{|\alpha|_1 \leq t_d} \left( \int_{Y(\mathcal{A}_n)} \int_D |D_y^\gamma D_x^\alpha \mathbf{P}^n q^*|^2 \rho_n(y) dx dy \right) \leq \|\mathbf{P}^n q^*\|_{\tilde{H}_{\text{mix}}}^2 \leq 2\|\mathbf{P}^n q^*\|_{\tilde{H}}^2 \end{aligned}$$

from which it follows that

$$\begin{aligned} \|u(q_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 &\leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 + \beta_n \|\hat{q}_n^*\|_{\tilde{H}_{\text{mix}}}^2 - \beta_n \|q_n^*\|_{\tilde{H}_{\text{mix}}}^2 \\ &\leq \|u(\hat{q}_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 + \beta_n \|\mathbf{P}^n q^*\|_{\mathcal{H}}^2. \end{aligned}$$

By Lemmas 3.5 and 2.2

$$\limsup_{n \rightarrow \infty} \|u(q_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 \leq \lim_{n \rightarrow \infty} \|u(\hat{q}_n^*) - \hat{u}^n\|_{\mathcal{H}_0^1}^2 + \beta_n \|\mathbf{P}^n q^*\|_{\mathcal{H}}^2 = \|u(q^*) - \hat{u}\|_{\mathcal{H}_0^1}^2,$$

which, together with the Banach Alaoglu Theorem, guarantees the existence of a subsequence  $u(q_{n_j}^*)$  converging weakly to some  $u_0 \in \mathcal{H}_0^1$ . Since feasible sets  $\{Q_{\text{ad}}^n\}_{n=1}^\infty$  form a nested sequence, all functions  $q_n^* \in Q_{\text{ad}}^n \subset Q_{\text{ad}}^1$ , which is weakly compact (Lemma 2.1). The sequence  $q_n^* \in Q_{\text{ad}}$  therefore has a subsequence,  $q_{n_j}^* \rightharpoonup q_0 \in Q_{\text{ad}}^1$  in  $\mathcal{H}$ . Additionally, since  $Q_{\text{ad}}^n$  is nested and the graph of  $u$  is weakly closed (Lemma 2.3) we have  $q_0 \in \cap_{n=1}^\infty Q_{\text{ad}}^n = Q_{\text{ad}}$  and  $u_0 = u(q_0)$ . Therefore

$$\begin{aligned} \|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1}^2 &= \lim_{j \rightarrow \infty} \langle u(q_{n_j}^*) - \hat{u}^{n_j}, u(q_0) - \hat{u} \rangle_{\mathcal{H}_0^1} \\ &\leq \liminf_{j \rightarrow \infty} \|u(q_{n_j}^*) - \hat{u}^{n_j}\|_{\mathcal{H}_0^1} \|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1} \end{aligned} \quad (14)$$

$$\begin{aligned} &\leq \limsup_{j \rightarrow \infty} \|u(q_{n_j}^*) - \hat{u}^{n_j}\|_{\mathcal{H}_0^1} \|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1} \\ &\leq \|u(q^*) - \hat{u}\|_{\mathcal{H}_0^1} \|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1}, \end{aligned} \quad (15)$$

which implies  $\|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1} \leq \|u(q^*) - \hat{u}\|_{\mathcal{H}_0^1}$  and hence  $q_0 \in Q_{\text{ad}}$  is a minimizer for  $(P)$ . Inequalities (14) and (15) further imply

$$\lim_{j \rightarrow \infty} \|u(q_{n_j}^*) - \hat{u}^{n_j}\|_{\mathcal{H}_0^1} = \|u(q_0) - \hat{u}\|_{\mathcal{H}_0^1},$$

which, together with the weak convergence  $u(q_{n_j}^*) - \hat{u}^{n_j} \rightharpoonup u(q_0) - \hat{u}$ , implies  $u(q_{n_j}^*) - \hat{u}^{n_j} \rightarrow u(q_0) - \hat{u}$  due to (12). In addition, the fact that  $\hat{u}^{n_j} \rightarrow \hat{u}$  implies that  $u(q_{n_j}) \rightarrow u(q_0)$ . Finally, this argument holds for any convergent subsequence of  $\{q_n^*\}$  and hence the Theorem is proved.  $\square$

## 4 The Finite Noise Problem

The immediate benefit of using  $\tilde{H}_{\text{mix}}$  as an approximate search space is that it imbeds continuously in  $L^\infty(D \times \Gamma^n)$ , regardless of the size of the stochastic dimension  $n$ . By virtue of the tensor product structure of  $\tilde{H}_{\text{mix}}(\Gamma^n)$  we may consider Sobolev regularity component-wise, which, in conjunction with the compact imbedding of  $H^1(\Gamma_i)$  in  $L^\infty(\Gamma_i)$ , gives rise to this property as the following lemma shows.

**Lemma 4.1.** *The space  $\tilde{H}_{\text{mix}}$  imbeds continuously in  $L^\infty(D \times \Gamma^n)$  for all  $n \in \mathbb{N}$ .*

*Proof.* For any fixed value  $y_0$  of the random component  $y$  and any multi-index  $\gamma \in \mathbb{N}^n$ , the function  $D_y^\gamma q(\cdot, y_0) \in H^{t_d}(D)$  whenever  $|\gamma|_\infty \leq s$ . Similarly, if both spatial variable  $x$  and all but the  $i^{\text{th}}$  component  $y_i$  of the stochastic variable  $y$  are fixed at  $x_0$  and  $y_0^1, \dots, y_0^{i-1}, y_0^{i+1}, \dots, y_0^n$  respectively, and  $\alpha \in \mathbb{N}^d$ ,  $\gamma_i^* := (\gamma_1, \dots, \gamma_{i-1}, 0, \gamma_{i+1}, \dots, \gamma_n) \in \mathbb{N}^n$  are multi-indices satisfying  $|\alpha|_1 \leq t_d$ ,  $|\gamma_i^*|_\infty \leq 1$ , then the mixed derivative  $D_x^\alpha D_{y_i^*}^{\gamma_i^*} q(x_0, y_0^1, \dots, y_0^{i-1}, \cdot, y_0^{i+1}, \dots, y_0^n) \in H^1(\Gamma_i) \hookrightarrow L^\infty(\Gamma_i)$ . Therefore, by repeated application of the 1-dimensional Sobolev Imbedding Theorem [1]

$$\begin{aligned} \|q\|_{L^\infty(D \times \Gamma)} &= \max_{x \in D, y \in \Gamma^n} |q(x, y)| = \max_{y \in \Gamma^n} \|q(\cdot, y)\|_{L^\infty(D)} \leq C \max_{(y_1, \dots, y_n) \in \Gamma^n} \|D_x^\alpha q(\cdot, y)\|_{H^1(D)} \\ &\leq C \max_{(y_1, \dots, y_{n-1}) \in \Gamma^{n-1}} \left( \sum_{|\alpha|_1 \leq t_d} \int_D \left( \max_{y_n \in \Gamma_n} |D_x^\alpha q(x, y_1, \dots, y_n)| \right)^2 dx \right)^{\frac{1}{2}} \\ &\leq CC_{\Gamma_n} \max_{(y_1, \dots, y_{n-1}) \in \Gamma^{n-1}} \left( \sum_{|\alpha|_1 \leq t_d} \sum_{\gamma_n=0}^1 \int_D \int_{\Gamma_n} |D_x^\alpha D_{y_n}^{\gamma_n} q(x, y_1, \dots, y_n)|^2 d\omega dx \right)^{\frac{1}{2}} \\ &\leq \dots \\ &\leq C \prod_{i=1}^n C_{\Gamma_i} \left( \sum_{|\alpha|_1 \leq t_d} \sum_{|\gamma|_\infty \leq 1} \int_D \int_{\Gamma^n} |D_x^\alpha D_y^\gamma q(x, y)|^2 \rho_n(y) dy dx \right)^{\frac{1}{2}} = \tilde{C}_n \|q\|_{\tilde{H}_{\text{mix}}} \end{aligned}$$

for some constant  $\tilde{C}_n > 0$ , independent of  $q$ , but possibly dependent on the total dimension  $d = d_p + n$ .  $\square$

### 4.1 Differentiability and Existence of Lagrange Multipliers

The Fréchet differentiability of the equality constraint  $e_n(q, u)$  follows directly from its continuity in  $q$  and  $u$ , since  $e_n(q, u)$  is affine linear in both arguments. Continuity in  $u$  is straightforward. For  $u, \tilde{u} \in \tilde{H}_0^1(D)$ ,

$$\|e_n(q, u - \tilde{u})\|_{\tilde{H}_0^1}^2 = \int_{\Gamma^n} \int_D q |\nabla(u - \tilde{u})|^2 dx \rho_n dy \leq q_{\max} \|u - \tilde{u}\|_{\tilde{H}_0^1}^2.$$

Continuity in the parameter  $q$  can now also be established, thanks to Lemma 4.1. Indeed,

$$\|e_n(q - \tilde{q}, u)\|_{\tilde{H}_0^1}^2 = \int_{\Gamma^n} \int_D |(q - \tilde{q}) \nabla u|^2 dx \rho_n dy \leq \|q\|_{L^\infty(D \times \Gamma)} \|u\|_{\tilde{H}_0^1}^2 \leq \tilde{C}_n^2 \|q\|_{\tilde{H}_{\text{mix}}}^2 \|u\|_{\tilde{H}_0^1}^2$$

for any  $q, \tilde{q} \in \tilde{H}_{\text{mix}}$ . A simple calculation then reveals that the first derivative of  $e_n$  in the direction  $(h, v) \in \tilde{H}_{\text{mix}} \times \tilde{H}_0^1$  is given by:

$$D_{(q,u)}[e_n(q, u)](h, v) = D_q[e_n(q, u)]h + D_u[e_n(q, u)]v \in \tilde{H}_0^1, \quad (16)$$

where the partial derivatives satisfy

$$\begin{aligned} \langle D_q[e_n(q, u)]h, \phi \rangle_{\tilde{H}_0^1} &= \int_{\Gamma^n} \int_D h \nabla u \cdot \nabla \phi dx \rho_n dy = \langle h \nabla u, \nabla \phi \rangle \quad \text{and} \\ \langle D_u[e_n(q, u)]v, \phi \rangle_{\tilde{H}_0^1} &= \int_{\Gamma^n} \int_D q \nabla v \cdot \nabla \phi dx \rho_n dy = \langle q \nabla v, \nabla \phi \rangle \quad \text{for all } \phi \in \tilde{H}_0^1. \end{aligned}$$

We can now derive more traditional, gradient-based first order necessary optimality conditions.

**Theorem 4.2** (Existence of Lagrange Multipliers). *Let  $(q^*, u^*)$  be a minimizer for problem  $(P^n)$ . Then there exists a unique Lagrange multiplier  $\lambda^* \in \tilde{H}_0^1$  for which the Lagrange functional  $L : \tilde{H}_{\text{mix}} \times \tilde{H}_0^1 \times \tilde{H}_0^1 \rightarrow \mathbb{R}$ , defined by*

$$L(q, u; \lambda) := J(q, u) + \langle \lambda, e_n(q, u) \rangle_{\tilde{H}_0^1}$$

*satisfies*

$$D_{(q,u)}[L(q^*, u^*; \lambda^*)](h, v) \geq 0 \quad \text{for all } (h, v) \in C(q^*) \times \tilde{H}_0^1, \quad (17)$$

*where*

$$C(q^*) = \{l(c - q^*) : c \in Q_{\text{ad}}, 0 \leq l \in \mathbb{R}\}.$$

*Particularly, the adjoint equation and complementary condition hold*

$$\langle q^* \nabla \lambda^*, \nabla \phi \rangle = -\langle u^* - \hat{u}^n, \phi \rangle_{\tilde{H}_0^1} \quad (18)$$

$$\beta \langle q^*, q - q^* \rangle_{\tilde{H}_{\text{mix}}} + \langle (q - q^*) \nabla u^*, \nabla \lambda^* \rangle \geq 0 \quad \text{for all } q \in Q_{\text{ad}}. \quad (19)$$

*Proof.* Let  $(q^*, u^*)$  be a minimizer of problem  $(P^n)$ . We show that  $(q^*, u^*)$  satisfies the regular point condition

$$D_{(q,u)}[e_n(q^*, u^*)](C(q^*) \times \tilde{H}_0^1) = \tilde{H}_0^1, \quad (20)$$

from which the existence of the Lagrange multiplier follows directly by [26]. In light of (16), this amounts to establishing the existence of solutions  $(h, v) \in C(q^*) \times \tilde{H}_0^1$  to the equation

$$D_q[e_n(q^*, u^*)]h + D_u[e_n(q^*, u^*)]v = w,$$

for arbitrary  $w \in \tilde{H}_0^1$ . Since  $0 \in C(q^*)$  and the finite noise elliptic equation

$$\int_{\Gamma^n} \int_D q \nabla v \cdot \nabla \phi dx \rho_n dy = \int_{\Gamma^n} \int_D \nabla w \cdot \nabla \phi dx \rho_n dy \quad \forall \phi \in \tilde{H}_0^1$$

is solvable for any  $w \in \tilde{H}_0^1$ , condition (20) is satisfied and hence there exists a Lagrange multiplier  $\lambda^* \in \tilde{H}_0^1$  such that (17) holds. More explicitly,

$$\begin{aligned} 0 &\leq \langle u^* - \hat{u}, v \rangle_{\tilde{H}_0^1} + \beta \langle q^*, h \rangle_{\tilde{H}_{\text{mix}}} + \langle D_q[e_n(q^*, u^*)]h + D_u[e_n(q^*, u^*)]v, \lambda^* \rangle_{\tilde{H}_0^1} \\ &= \langle u^* - \hat{u}, v \rangle_{\tilde{H}_0^1} + \beta \langle q^*, h \rangle_{\tilde{H}_{\text{mix}}} + \langle h \nabla u^*, \nabla \lambda^* \rangle + \langle q^* \nabla v, \nabla \lambda^* \rangle \end{aligned} \quad (21)$$

for all  $(h, v) \in C(q^*) \times \tilde{H}_0^1$ . In particular, if  $h = 0$ , we obtain

$$\langle q^* \nabla \lambda^*, \nabla v \rangle = -\langle u^* - \hat{u}, v \rangle_{\tilde{H}_0^1} \quad \text{for all } v \in \tilde{H}_0^1,$$

which yields the adjoint equation (18). The uniqueness of  $\lambda^*$  now follows directly from the uniqueness of the solution to the elliptic equation (18). Finally, setting  $v = 0$  and  $h = q - q^*$  in (21) for any  $q \in Q_{\text{ad}}$  yields the complementary condition (19)

$$\beta \langle q^*, q - q^* \rangle_{\tilde{H}_{\text{mix}}} + \langle (q - q^*) \nabla u^*, \nabla \lambda^* \rangle \geq 0 \quad \text{for all } q \in Q_{\text{ad}}.$$

□

## 5 An Augmented Lagrangian Algorithm

With the availability of derivative information, the finite noise problem  $(P^n)$  can now be solved by more conventional optimization algorithms. We make use of the augmented Lagrangian method, an iterative approach that may be viewed as a modified penalty method. The quadratic penalty method avoids explicit enforcement of the equality constraint  $e_n(q, u) = 0$  by incorporating an additional term, that penalizes violations of the constraint, into the cost functional. For example in  $(P^n)$ , this could require solving a series of sub-problems of the form

$$\min_{(q, u) \in Q_{\text{ad}} \times \tilde{H}_0^1} \frac{1}{2} \|u - \hat{u}\|_{\tilde{H}_0^1}^2 + \frac{\beta}{2} \|q\|_{\tilde{H}_{\text{mix}}}^2 + \frac{c_k}{2} \|e_n(q, u)\|_{\tilde{H}_0^1}^2, \quad (22)$$

where the sequence  $\{c_k\}_{k=0}^\infty$  increases steadily as  $k \rightarrow \infty$ . In fact, the convergence of this class of methods requires  $\lim_{k \rightarrow \infty} c_k = \infty$ , leading to a progressive deterioration in the conditioning of the sub-problem.

The augmented Lagrangian method avoids this conditioning issue by instead solving the sequence of problems

$$\min_{(q, u) \in Q_{\text{ad}} \times \tilde{H}_0^1} L_{c_k}(q, u, \lambda^k), \quad (P_{\text{aux}})$$

where  $\{c_k\}_{k=0}^\infty$  is a non-decreasing sequence of positive numbers and the augmented Lagrangian functional,  $L_{c_k} : \tilde{H}_{\text{mix}} \times \tilde{H}_0^1 \times \tilde{H}_0^1 \rightarrow \mathbb{R}$ , is given by

$$L_{c_k}(q, u, \lambda^k) = \frac{1}{2} \|u - \hat{u}\|_{\tilde{H}_0^1}^2 + \frac{\beta}{2} \|q\|_{\tilde{H}_{\text{mix}}}^2 + \langle \lambda^k, e_n(q, u) \rangle_{\tilde{H}_0^1} + \frac{c_k}{2} \|e_n(q, u)\|_{\tilde{H}_0^1}^2.$$

The function  $\lambda^k \in \tilde{H}_0^1$  is an approximation of the Lagrange multiplier defined in (18) and is updated via  $\lambda^{k+1} = \lambda^k + c_k e_n(q^k, u^k)$ , where  $(q^k, u^k)$  minimizes  $(P_{\text{aux}})$ . More explicitly,

**Input** :  $\hat{u}$   
**Output**:  $q$   
**1** Choose  $\lambda^0 \in H_0^1(D)$ , and non-decreasing sequence  $\{c_k\}$  with  $c_0 > 0$ ;  
**2** Set  $k = 0$ ;  
**3** **while** *not converged* **do**  
**4**     Obtain minimizers  $(q^k, u^k)$  by solving the auxiliary problem  $(P_{\text{aux}})$ ;  
**5**     Set  $\lambda^{k+1} := \lambda^k + c_k e_n(q^k, u^k)$ ;  
**6**     Set  $k = k + 1$  and test for convergence;  
**7** **end**

**Algorithm 1:** The Augmented Lagrangian Algorithm

This algorithm, developed in [16, 29], has been used extensively for deterministic parameter identification- and control problems in elliptic systems [17, 18, 21]. Unlike for penalty methods, the sequence  $\{c_k\}_{k=0}^\infty$  is not required to grow without bound to guarantee convergence.

It was shown in [18] and [21] (Theorems 2.4, 2.5, and subsequent remarks) that the iterates  $(q^k, u^k, \lambda^k)$  computed by Algorithm 1 converge to the minimizers  $(q^*, u^*, \lambda^*)$  of  $(P^n)$ , under the following second-order sufficient optimality condition:

**Assumption 5.1.** Assume there exists a constant  $\tau = \tau(\beta) > 0$  so that

$$D_{(q,u)}^2[L(q^*, u^*, \lambda^*)](h, v)^2 \geq \tau(\|h\|_{\tilde{H}_{\text{mix}}}^2 + \|v\|_{\tilde{H}_0^1}^2) \quad \text{for all } (h, v) \in \tilde{H}_{\text{mix}} \times \tilde{H}_0^1.$$

The original convergence proof, formulated in a general Hilbert space setting, carries over directly to our problem. We refer the interested reader to the cited references. Moreover, the cost functional  $L_{c_k}$  appearing in the auxiliary problem  $(P_{\text{aux}})$  is quadratic in  $q$  for fixed  $u$  and  $\lambda$  and quadratic in  $u$  for fixed  $q$  and  $\lambda$ , suggesting the use of sequential splitting methods to speed up the solution of the auxiliary subproblem. To wit, the subproblem  $(P_{\text{aux}})$  in Algorithm 1 is replaced with the sequence: Solve

$$\min_{q \in Q_{\text{ad}}} L_{c_k}(q, u_{n,k}^*, \lambda_{n,k}^*). \quad (P_{\text{aux}}^q)$$

for  $q_{n,k}^*$ , then obtain  $u_{n,k+1}^*$  by solving the minimization problem

$$\min_{u \in H_0^1} L_{c_k}(q_{n,k+1}^*, u, \lambda_{n,k}^*). \quad (P_{\text{aux}}^u)$$

- 1 Choose  $\lambda_{n,0} \in h(D)$ , and non-decreasing sequence  $\{c_k\}$  with  $c_0 > 0$ ;
- 2 Set  $k = 0$  ;
- 3 **while not converged do**
- 4     Solve the auxiliary problem sequentially, i.e. for iterates  $q_{n,k+1}^*$  and  $u_{n,k+1}^*$ ;
- 5     Get  $q_{n,k+1}^*$  by solving problem  $(P_{\text{aux}}^q)$  (using current values of  $u_{n,k}^*$  and  $\lambda_{n,k}^*$ );
- 6     Get  $u_{n,k+1}^*$  by solving problem  $(P_{\text{aux}}^u)$  (using current values of  $q_{n,k+1}^*$  and  $\lambda_{n,k}^*$ );
- 7     Set  $\lambda_{n,k+1}^* := \lambda_{n,k}^* + c_k e_n(q_{n,k+1}^*, u_{n,k+1}^*)$ ;
- 8     Set  $k = k + 1$  and test for convergence.
- 9 **end**

**Algorithm 2:** The Augmented Lagrangian Algorithm with Sequential Splitting

We consider the auxiliary sub-problems  $P_{\text{aux}}^q$  and  $P_{\text{aux}}^u$  in more detail. The unconstrained minimizer  $u_{n,k+1}^*$  of  $P_{\text{aux}}^u$  can be computed simply by solving the first order optimality system  $D_u[L_{c_k}(q, u, \lambda)](v) = 0$  for all  $v \in \tilde{H}_0^1$  and fixed  $q \in Q_{\text{ad}}, \lambda \in \tilde{H}_0^1$ , where

$$\begin{aligned} 0 &= D_u[L_{c_k}(q, u, \lambda)](v) \\ &= \langle u - \hat{u}, v \rangle_{\tilde{H}_0^1} + \langle \lambda, D_u[e_n(q, u)](v) \rangle_{\tilde{H}_0^1} + c_k \langle e_n(q, u), D_u[e_n(q, u)](v) \rangle_{\tilde{H}_0^1} \\ &= \langle u - \hat{u}, v \rangle_{\tilde{H}_0^1} + \langle q \nabla \lambda, \nabla v \rangle + c_k \langle q \nabla e_n(q, u), \nabla v \rangle \\ &= \langle \nabla u + c_k q \nabla e_n(q, u), \nabla v \rangle - \langle \hat{u} - q \nabla \lambda, \nabla v \rangle. \end{aligned} \quad (23)$$

The first order optimality system for  $P_{\text{aux}}^q$  if  $q \in \text{int}(Q_{\text{ad}})$  amounts to setting  $D_q[L_{c_k}(q, u, \lambda)](h) = 0$  for all  $h \in \tilde{H}_{\text{mix}}$ . More specifically,

$$\begin{aligned} 0 &= D_q[L_{c_k}(q, u, \lambda)](h) \\ &= \beta \langle q, h \rangle_{\tilde{H}_{\text{mix}}} + \langle \lambda, D_q[e_n(q, u)](h) \rangle_{\tilde{H}_0^1} + c_k \langle e_n(q, u), D_q[e_n(q, u)](h) \rangle_{\tilde{H}_0^1} \\ &= \beta \langle q, h \rangle_{\tilde{H}_{\text{mix}}} + \langle h \nabla \lambda, \nabla u \rangle + c_k \langle h \nabla e_n(q, u), \nabla u \rangle. \end{aligned} \quad (24)$$



## 6 Numerical Discretization

This section details the numerical discretization of the augmented Lagrangian method (Algorithm 2) outlined in the previous section. We approximate the parameter-  $q$ , state-  $u$ , and adjoint random fields  $\lambda$  spatially by means of piecewise polynomial basis functions related to finite element meshes of the spatial domain  $D$ . For the deterministic parameter identification problem, it was observed in [17] that using a coarser mesh for the parameter space than for the state space amounts to an implicit regularization. For our numerical experiments, we therefore base our approximation of  $q$  on a coarser triangulation  $\mathcal{T}_q$  of  $D$  with associated finite element space  $V_q = \text{span}\{\phi_1^q, \dots, \phi_{M_q}^q\}$ , while estimating  $u$  and  $\lambda$  based on the finer grid  $\mathcal{T}_u$ , in our case a uniform refinement of  $\mathcal{T}_q$ , with associated subspace  $V_u = \text{span}\{\phi_1^u, \dots, \phi_{M_u}^u\}$ . The spatial approximation  $v^{M_u} \in V_u \otimes L^2(\Omega)$  of  $v \in \mathcal{H}_0^1$  can be written explicitly as

$$v^{M_u}(x, \omega) := \sum_{i=1}^{M_u} v(x_i, \omega) \phi_i^u(x).$$

Estimates of associated spatial inner products can be also be computed using the mass- and stiffness matrices defined component-wise by

$$A^u := \left[ \int_D \phi_{i_1}^u(x) \phi_{i_2}^u(x) dx \right]_{i_1, i_2=1}^{M_u} \quad \text{and} \quad A_x^u := \left[ \int_D \nabla \phi_{i_1}^u(x) \cdot \nabla \phi_{i_2}^u(x) dx \right]_{i_1, i_2=1}^{M_u}$$

respectively. Similar expressions hold for the spatial approximations  $h^{M_q} \in V_q \otimes L^2(\Omega)$  of random fields  $h \in \mathcal{H}$  and for the mass- and stiffness matrices  $A^q$  and  $A_x^q$  on  $V_q$ , although we assume here that homogeneous Dirichlet boundary conditions are incorporated into the construction of  $A_x^u$ , rendering it invertible, while no such conditions are imposed on  $A_x^q$ .

### 6.1 Karhunen-Loève Expansion of the Data

In order to reduce our variational problem  $(P)$  to its ‘finite noise’ approximation  $(P^n)$ , we must first approximate the truncated KL expansion of the measured data  $\hat{u} \in \mathcal{H}_0^1$ , which in turn requires the spectral decomposition of the compact covariance operator  $\mathcal{C}_{\hat{u}} : H_0^1(D) \rightarrow H_0^1(D)$ , defined in terms of its covariance kernel

$$C_{\hat{u}}(x, x') = \mathbb{E}[(\hat{u}(x') - u_0(x'))(\hat{u}(x) - u_0(x))]$$

$$v \in H_0^1(D) \mapsto (\mathcal{C}_{\hat{u}} v)(x') = \int_D \nabla_x C_{\hat{u}}(x, x') \cdot \nabla v(x) dx \in H_0^1(D),$$

where  $u_0(x) := \mathbb{E}[\hat{u}(x, \cdot)]$ . In practice,  $\hat{u}$  commonly occurs in the form of an data matrix  $\hat{\mathbf{U}} = [\hat{u}_{i,j}]$ , where  $\hat{u}_{i,j} = \hat{u}(x_i, \omega_j)$  denotes the  $j^{\text{th}}$  random sample of the field obtained at spatial location  $x_i$  for  $j = 1, \dots, N_{\text{sample}}$ . We assume here that this data is either sampled at the vertices  $x_i$  of the grid  $\mathcal{T}_u$ , or that it is interpolated, using splines for example, so that  $\hat{\mathbf{U}}$  is of size  $M_u$  by  $N_{\text{sample}}$ . Let the sample mean  $\mathbf{m} = [m_1, \dots, m_{M_u}]^T$  and covariance matrix  $\Sigma = [\sigma_{i_1, i_2}]_{i_1, i_2=1}^{M_u}$  be defined componentwise by

$$m_i := \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} \hat{u}(x_i, \omega_j), \text{ and}$$

$$\sigma_{i_1, i_2} := \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} (\hat{u}(x_{i_1}, \omega_j) - m_{i_1})(\hat{u}(x_{i_2}, \omega_j) - m_{i_2}),$$

respectively. The sample mean  $\hat{u}_0^{M_u}$  and covariance  $C_{\hat{u}}^{M_u}$  of a finite element representation  $\hat{u}^{M_u}$  of  $\hat{u}$  then take the form

$$\begin{aligned}\hat{u}_0^{M_u}(x) &= \sum_{i=1}^{M_u} m_i \phi_i^u(x) \quad \text{and} \\ C_{\hat{u}}^{M_u}(x, x') &= \sum_{i_1, i_2} \sigma_{i_1, i_2} \phi_{i_1}^u(x) \phi_{i_2}^u(x'),\end{aligned}$$

respectively. This allows us to form the finite element approximation  $\mathcal{C}_{\hat{u}}^{M_u} : V_u \rightarrow V_u$  of the covariance operator by letting

$$\begin{aligned}(\mathcal{C}_{\hat{u}}^{M_u} v)(x') &= \int_D C_{\hat{u}}^{M_u}(x, x') v(x) dx \\ &= \sum_{i_1=1}^{M_u} v(x_{i_1}) \phi_{i_1}(x') \left( \sum_{i_2=1}^{M_u} \sigma_{i_1, i_2} \int_D \nabla \phi_{i_1}(x) \cdot \nabla \phi_{i_2}(x) dx \right)\end{aligned}$$

for any element  $v \in V_u$ . The operation  $\mathcal{C}_{\hat{u}}^{M_u} v$  can also be expressed in terms of the spatial coordinatization  $\mathbf{v} = [v(x_1), \dots, v(x_{M_u})]^T$  of  $v$  as the matrix-vector product  $\Sigma A_x^u \mathbf{v}$  and hence the spectral decomposition of  $\mathcal{C}_{\hat{u}}^{M_u}$  amounts to finding the eigenpairs  $(\nu, \mathbf{b})$  so that  $\Sigma A_x^u \mathbf{b} = \nu \mathbf{b}$ , or equivalently the generalized eigenvalue problem  $A_x^u \Sigma A_x^u \mathbf{b} = \nu A_x^u \mathbf{b}$ . By virtue of the positive semi-definiteness of the discretized covariance operator  $\mathcal{C}_{\hat{u}}^{M_u}$  the eigenvectors  $\mathbf{b}$  are orthogonal, so that the associated eigen-decomposition takes the form  $\Sigma A_x^u = B D^\nu B^T$  with  $D^\nu$  diagonal and  $B$  unitary. The truncated KL expansion amounts to a projection of the data onto the eigenspace associated with the largest  $n$  eigenvalues. The compactness and semi-positive definiteness of the operator  $\mathcal{C}_{\hat{u}}$  ensure that its spectrum is countable with an accumulation point at 0, allowing us to determine a suitable truncation level  $n$  by estimating the rate of decay of the eigenvalues. Since  $\mathcal{C}_{\hat{u}}^{M_u}$  only has finite rank, however, this criterion is subject to the level of spatial discretization  $M_u$ , i.e. we require  $n \leq M_u$ . The truncated, discretized KL expansion  $\hat{u}^{n, M_u}$  of the field  $\hat{u}$  now takes the form

$$\hat{u}^{n, M_u}(x, \omega) = \hat{u}_0^{M_u}(x) + \sum_{k=1}^n \sqrt{\nu_k} b_k^{M_u}(x) Y_k(\omega) \quad \text{for } \omega \in \Omega,$$

where  $Y(\omega) = [Y_1(\omega), \dots, Y_n(\omega)]^T$  is a random vector whose joint density function can be estimated from samples obtained by projecting the centered data matrix onto the subspace spanned by the dominant  $n$  eigenvectors. Indeed, let  $B_n$  be the matrix consisting of the first  $n$  columns of  $B$  and  $D_n^\nu = \text{diag}(\nu_1, \dots, \nu_n)$ . Then

$$\begin{aligned}Y_k(\omega_j) &= \frac{1}{\sqrt{\nu_k}} \int_D \nabla \left( \hat{u}^{n, M_u}(x, \omega_j) - \hat{u}_0^{M_u}(x) \right) \cdot \nabla b_k^{M_u}(x) dx \\ &= \sum_{i_1, i_2=1}^{M_u} \frac{1}{\sqrt{\nu_k}} (\hat{u}(x_{i_1}, \omega_j) - \hat{u}_0(x_{i_1})) b_k(x_{i_2}) \int_D \nabla \phi_{i_1}^u(x) \cdot \nabla \phi_{i_2}^u(x) dx\end{aligned}$$

for  $k = 1, \dots, n$ , so that  $Y(\omega_j) = (D_n^\nu)^{-\frac{1}{2}} B_n^T A_x^u \left( \hat{\mathbf{U}}(:, j) - \mathbf{m} \right)$  for  $j = 1, \dots, N_{\text{sample}}$ . It is from these samples that the joint density function  $\rho_n$  can be estimated. The KL expansion discussed in this paper differs slightly from the usual approach [33], in that we are defining the covariance operator on the Hilbert space  $H_0^1(D)$  instead of on  $L^2(D)$ , to ensure convergence of the projection in the  $\mathcal{H}_0^1$  norm. In practice, this choice of the norm doesn't make a significant difference in computations.

The estimation of multidimensional density functions is a highly non-trivial problem in general and an active field of current statistical research, well beyond the scope of this paper. The reader is referred to the books [35, 20], as well as the survey article [34], for a more exhaustive treatment of the subject. The random vectors encountered in Section 7 are only of moderate size and we either assume to know their joint densities or make use of kernel density estimators to approximate them empirically.

## 6.2 Discretization in the Stochastic Component

The choice of the type of nodal basis used to discretize the state equation ( $P^n$ ) or the adjoint system (18) depends on the smoothness of the fields  $u$  and  $\lambda$  as functions of  $y$ . Under certain smoothness conditions on the parameter  $q(x, y)$ , which are readily satisfied if  $q$  is written in terms of its KL expansion, the model output  $u(x, y)$  can be shown to be analytic in  $y$ , warranting the use of global interpolating basis functions such as Lagrange polynomials [2]. In our case  $q(x, y)$  is written in terms of the random variables in the KL expansion of the measured data  $\hat{u}$  and hence such smoothness conditions may no longer hold. Consequently, neither the model output  $u$ , nor the Lagrange multiplier  $\lambda$ , characterized by the adjoint equation, are guaranteed to exhibit the requisite smoothness as functions of  $y$  to allow for their approximation by a global polynomial basis. Here we make use of an interpolating basis of piecewise smooth, multi-linear hat functions.

Assume, without loss, of generality that the stochastic domain  $\Gamma^n = [0, 1]^n$ . While much is known about interpolation formulas on one-dimensional domains, the problem of computing efficient and accurate multi-dimensional interpolants remains a challenge. Sparse grid methods [7, 15, 28, 36] efficiently combine one-dimensional interpolation schemes to obtain accurate interpolants in higher dimensions with only a moderate number of grid points. Suppose  $\Gamma^n$  is subdivided along each dimension into one-dimensional grids  $X^{l_t}$ ,  $t = 1, 2, \dots, n$  of equally spaced points, where the multi-index  $\mathbf{l} = (l_1, \dots, l_n) \in \mathbb{N}^n$  denotes the level of refinement in each direction. In particular, each grid  $X^{l_t}$  consists of nodes  $\{y_{l_t, j_t}\}_{j_t=0}^{m^{l_t}}$ , where

$$m^{l_t} = \begin{cases} 1, & \text{if } l_t = 1 \\ 2^{l_t}, & \text{if } l_t > 1 \end{cases} \quad \text{and} \quad y_{l_t, j_t} = \begin{cases} 0.5, & \text{if } l_t = 1, j_t = 1 \\ 2^{-l_t} j_t, & \text{if } l_t > 1, \text{ for } j_t = 0, 1, \dots, m^{l_t} \end{cases}.$$

For convenience, we define  $m^{\mathbf{l}} := (m^{l_1}, \dots, m^{l_n})$  and take  $j \leq m^{\mathbf{l}}$  to mean  $j_t \leq m^{l_t}$  for each  $t = 1, \dots, n$ . The full tensor product grid  $X^{\mathbf{l}}$  on  $\Gamma^n$ , given by

$$X^{\mathbf{l}} := X^{l_1} \times \dots \times X^{l_n},$$

thus consists of the points  $\{y_{\mathbf{l}, j}\}_{j \leq m^{\mathbf{l}}}$ . Let  $\{\psi_{l_t, j_t}\}_{j_t=0}^{m^{l_t}}$  denote a set of one-dimensional, nodal interpolating basis functions centered at the grid points  $\{y_{l_t, j_t}\}_{j_t=0}^{m^{l_t}}$  of each one-dimensional grid  $X^{l_t}$ ,  $t = 1, \dots, n$ . We use bases of one-dimensional piecewise linear hat functions, defined for any point  $y \in [0, 1]$  by  $\psi_{l_t, j_t}(y) := 1$  when  $l_t = 1$  and

$$\psi_{l_t, j_t}(y) := \psi\left(m^{l_t} \left(y - \frac{j_t}{m^{l_t}}\right)\right), \quad \psi(z) := \begin{cases} 1 - |z|, & \text{if } -1 \leq z \leq 1 \\ 0, & \text{otherwise} \end{cases},$$

when  $l_t > 1$ . A basis function  $\psi_{\mathbf{l}, j}$  centered at a node  $y_{\mathbf{l}, j} = (y_{l_1, j_1}, \dots, y_{l_n, j_n})$  in the multi-dimensional grid  $X^{\mathbf{l}} = X^{l_1} \times \dots \times X^{l_n} \subset [0, 1]^n$  can then be obtained by taking the product of the appropriate univariate nodal basis functions, i.e. for any  $y = (y_1, \dots, y_n) \in [0, 1]^n$ ,

$$\psi_{\mathbf{l}, j}(y) = \psi_{l_1, j_1} \otimes \dots \otimes \psi_{l_n, j_n}(y) := \prod_{t=1}^n \psi_{l_t, j_t}(y_t).$$

Note that the one-dimensional grids are nested, i.e.  $X^0 \subset X^1 \subset \dots \subset X^{l_t}$  for any  $l_t \in \mathbb{N}$ . As a result, the subspaces spanned by one-dimensional interpolating basis functions are also nested and hence it is relatively straightforward to compare the accuracy of one-dimensional grids with various refinement levels  $l_t$ . A multi-dimensional interpolation formula with refinement level  $L$  in each direction can be obtained by combining the one-dimensional interpolation formulas

$$U^L(v) = \sum_{j_t=0}^{m^L} v(y_{l_t, j_t}) \psi_{l_t, j_t}$$

to form the full tensor multi-variate interpolant

$$(U^L \otimes \dots \otimes U^L)(v) = \sum_{j \leq m^L} v(y_{l, j}) \psi_{l, j}.$$

The number of grid points needed to construct this interpolant is  $(m^L)^n$ , which scales exponentially as the dimension  $n$  of the space increases.

The sparse grid interpolant  $A^L(v)$  with interpolation level  $L \geq 0$  is constructed from linear combinations of lower order full tensor interpolants as follows

$$A^L(v) = \sum_{1 \leq |l|_1 \leq L+n-1} (-1)^{N-|l|_1} \binom{n-1}{L-|l|_1} (U^{l_1} \otimes \dots \otimes U^{l_n})(v). \quad (25)$$

Through cancellation, the effective number of grid points required is much lower than that of the full tensor product, while its accuracy is only marginally worse.

In practice, formula (25) is not used directly to construct interpolants. Instead, higher order interpolants are constructed recursively from lower order ones by adding corrections on the appropriately refined grid. This is achieved through the use of hierarchical basis functions, defined for every level  $l = (l_1, \dots, l_n)$  to be the span  $W^l(\Gamma^n) = \text{span}\{\psi_{l, j} : j \in J_l\}$ , where

$$J_l = \left\{ j \in \mathbb{N}^n : j_t = \begin{cases} 1/2 & \text{if } l_t = 1, \\ 0 \text{ or } 1 & \text{if } l_t = 2, \\ \text{an odd number in } \{1, \dots, m^{l_t} - 1\} & \text{if } l_t \geq 3 \end{cases} \right\}.$$

Indeed, it can be shown (see [10]) that  $A^1(v) = (U^1 \otimes \dots \otimes U^1)(v)$ , while for any  $L > 1$

$$A^L(v) = A^{L-1}(v) + \Delta A^L(v),$$

where

$$\Delta A^L(v) = \sum_{|l|_1=L+n-1} \sum_{j \in J_l} \left[ v(y_{l, j}) - A^{L-1}(v)(y_{l, j}) \right] \cdot \psi_{l, j}(y).$$

The coefficients  $v_z(y_{l, j}) = v(y_{l, j}) - A^{L-1}(v)(y_{l, j})$  appearing in the update  $\Delta A^L$ , also known as hierarchical surpluses, represent the discrepancy between the function  $v$  and the  $L-1$  level interpolant  $A^{L-1}(v)$  at the new gridpoints. Hierarchical surpluses provide useful *a posteriori* error estimates that can readily be employed by an adaptive scheme to identify the regions where the grid should be refined [10, 24, 25]. Unfortunately, it is difficult to incorporate adaptive approximation seamlessly into these high-dimensional gradient-based optimization methods. Since the functions  $q_k, u_k$  and  $\lambda_k$  are changing at each iteration of the optimization algorithm, the adaptive refinement scheme would have to be adjusted throughout the duration of the algorithm. This can be costly, especially in light of the fact that the relevant bilinear- and trilinear forms would have to be updated after each adaptive refinement or coarsening.

For the sake of notational expediency, we let  $j = 1, \dots, N$  be an enumeration of the sparse grid points, i.e.

$$\{y_j\}_{j=1}^N = \{y_{l,j} : 1 \leq |l|_1 \leq L + n - 1, j \in J_l\},$$

so that the stochastic sparse grid interpolant  $v^N(x, y)$  of  $v \in \tilde{H}_0^1(D)$  takes the form

$$v^N(x, y) = \sum_{j=1}^N v_z(x, y_j) \psi_j(y),$$

while the full approximation of  $v$  is given by

$$v^{M_u, N}(x, y) = \sum_{i=1}^{M_u} \sum_{j=1}^N v_z(x_i, y_j) \phi_i^u(x) \psi_j(y)$$

The function values  $v(x_i, y_j)$  can be related to the hierarchical surpluses  $v_z(x_i, y_j)$  by means of a linear, invertible transformation.

### 6.3 The Discretized Optimization Problem

To approximate the inner products and bilinear forms appearing in optimization Algorithm 2, we require the deterministic bilinear forms introduced earlier, the  $\rho$ -weighted stochastic bilinear forms  $S_\rho$  and  $S_\rho^{\text{mix}}$ , and the stochastic trilinear form  $T_\rho$ , defined componentwise as follows

$$\begin{aligned} S_\rho &= \left[ \int_{\Gamma^n} \psi_{i_1}(y) \psi_{i_2}(y) \rho_n(y) dy \right]_{i_1, i_2=1}^N, \\ S_\rho^{\text{mix}} &= \left[ \sum_{|\gamma|_\infty \leq s} \int_{\Gamma^n} D_y^\gamma \psi_{i_1}(y) D_y^\gamma \psi_{i_2}(y) \rho_n(y) dy \right]_{i_1, i_2=1}^N, \text{ and} \\ T_\rho &= \left[ \int_{\Gamma^n} D_y^\gamma \psi_{i_1}(y) D_y^\gamma \psi_{i_2}(y) \psi_{i_3}(y) \rho_n(y) dy \right]_{i_1, i_2, i_3=1}^N. \end{aligned}$$

The evaluation of these multi-dimensional integrals for any given density function  $\rho$  is a challenging task in general, although they can be computed offline. Note that, whereas each basis function  $\psi_j(y)$  can be written as the product of appropriate one-dimensional basis functions, the  $\rho$  cannot in general be decomposed as the product of its marginals, thus preventing the effective decoupling of these integrals into products of simpler ones.

For any function  $v \in \tilde{H}_0^1(D)$ , we define  $\mathbf{v}^z := [\mathbf{v}_1^z, \dots, \mathbf{v}_N^z]^T$  to be the vector of hierarchical surpluses where  $\mathbf{v}_j^z = [v_z(x_1, y_j), \dots, v_z(x_{M_u}, y_j)]^T$  are the surpluses corresponding to the sparse grid node  $y_j$ . Let a similar definition hold for functions  $h \in \tilde{H}_{\text{mix}}^1(D)$ . The  $\tilde{H}_0^1$ -inner product of approximations  $v^{M_u, N}$  and  $w^{M_u, N}$  then take the form

$$\begin{aligned} &\langle v^{M_u, N}, w^{M_u, N} \rangle_{\tilde{H}_0^1} \\ &= \sum_{i_1, i_2=1}^{M_u} \sum_{j_1, j_2=1}^N v_z(x_{i_1}, y_{j_1}) w_z(x_{i_2}, y_{j_2}) \left( \int_{\Gamma^n} \psi_{j_1} \psi_{j_2} \rho dy \right) \left( \int_D \nabla \phi_{i_1}^u \cdot \nabla \phi_{i_2}^u dx \right) \\ &= \sum_{j_1, j_2=1}^N (\mathbf{v}_{j_2}^z)^T A_x^u \mathbf{w}_{j_1}^z = (\mathbf{v}^z)^T (S_\rho \otimes A_x^u) \mathbf{w}^z. \end{aligned}$$

Similarly,

$$\begin{aligned}\langle v^{M_u, N}, w^{M_u, N} \rangle_{\tilde{L}^2} &= (\mathbf{v}^z)^T (S_\rho \otimes A^u) \mathbf{w}^z, \quad \text{and} \\ \langle h^{M_q, N}, k^{M_q, N} \rangle_{\tilde{H}_{\text{mix}}} &= (\mathbf{h}^z)^T (S_\rho^{\text{mix}} \otimes A_x^q) \mathbf{k}^z,\end{aligned}$$

for any two functions  $h, k \in \tilde{H}_{\text{mix}}(D)$ . The discretized  $q$ -weighted bilinear form  $\langle q^{M_q, N} \nabla v^{M_u, N}, \nabla w^{M_u, N} \rangle$  on the other hand requires the use of the weighted trilinear form  $T_\rho$ . Indeed

$$\begin{aligned}\langle q^{M_q, N} \nabla u^{M_u, N}, \nabla v^{M_u, N} \rangle &= \int_{\Gamma^n} \int_D q^{M_q, N} \left( \nabla u^{M_u, N} \cdot \nabla v^{M_u, N} \right) \rho \, dx \, dy \\ &= \sum_{i_1, i_2=1}^{M_u} \sum_{j_1, j_2=1}^N u_z(x_{i_1}, y_{j_1}) v_z(x_{i_2}, y_{j_2}) \int_{\Gamma^n} \int_D q^{M_q, N} \nabla \phi_{i_1}^u \cdot \nabla \phi_{i_2}^u \psi_{j_1} \psi_{j_2} \rho \, dx \, dy \\ &= (\mathbf{u}^z)^T S_{\rho, q} \mathbf{v}^z,\end{aligned}$$

where  $S_{\rho, q}$  is defined componentwise as

$$S_{\rho, q} := \left[ \sum_{i=1}^{M_q} \sum_{j=1}^N q_z(x_i, y_j) \left( \int_{\Gamma^n} \psi_j \psi_{j_1} \psi_{j_2} \rho \, dy \right) \left( \int_D \phi_i^q \nabla \phi_{i_1}^u \cdot \nabla \phi_{i_2}^u \, dx \right) \right]_{\substack{i_1, i_2=1, \dots, M_u \\ j_1, j_2=1, \dots, N}}.$$

Alternatively,

$$\langle q^{M_q, N} \nabla u^{M_u, N}, \nabla v^{M_u, N} \rangle = (\mathbf{q}^z)^T (S_{\rho, u}) \mathbf{v}^z,$$

where

$$S_{\rho, u} := \left[ \sum_{i=1}^{M_u} \sum_{j=1}^N u_z(x_i, y_j) \left( \int_{\Gamma^n} \psi_j \psi_{j_1} \psi_{j_2} \rho \, dy \right) \left( \int_D \phi_i^q \nabla \phi_{i_1}^u \cdot \nabla \phi_{i_2}^u \, dx \right) \right]_{\substack{i_1, i_2=1, \dots, M_u \\ j_1, j_2=1, \dots, N}}.$$

In our numerical calculations, we approximate the sample paths of the equality constraint  $e \in \tilde{H}_0^1(D)$  as solutions to the spatially discretized Poisson problems

$$\int_D \nabla e_j^{M_u} \cdot \nabla \phi_i^u \, dx = \int_D q^{M_q, N}(\cdot, y_j) \nabla u^{M_u, N}(\cdot, y_j) \cdot \nabla \phi_i^u \, dx - \int_D f \phi_i^u \, dx, \quad (26)$$

$i = 1, \dots, M^u$ , or equivalently

$$\mathbf{e}_j = \mathbf{e}_j(\mathbf{q}, \mathbf{u}) - \mathbf{e}_j(\mathbf{f}),$$

for each  $j = 1, \dots, N$ , where

$$\begin{aligned}\mathbf{e}_j(\mathbf{q}, \mathbf{u}) &= (A_x^u)^{-1} H^j(\mathbf{q}, \mathbf{u}), \quad \mathbf{e}_j(\mathbf{f}) = (A_x^u)^{-1} \mathbf{f}, \quad \text{and} \\ H^j(\mathbf{q}, \mathbf{u}) &= \left[ \sum_{i_1=1}^{M^q} \sum_{i_2=1}^{M_u} q(x_{i_1}, y_j) u(x_{i_2}, y_j) \int_D \phi_{i_1}^q \nabla \phi_{i_2}^u \cdot \phi_i^u \, dx \right]_{i=1}^{M^u}.\end{aligned}$$

The vector  $\mathbf{e} = [\mathbf{e}_1, \dots, \mathbf{e}_N]^T$  of sample paths  $\mathbf{e}_j = [e^{M_u}(x_1, y_j), \dots, e^{M_u}(x_{M_u}, y_j)]^T$  for  $j = 1, \dots, N$ , can now be converted to the appropriate set of hierarchical surpluses  $\mathbf{e}^z$  through a standard linear transformation. Note that the system solves required to evaluate  $\mathbf{e}_j$  involve the same coefficient matrix, but with multiple right hand sides, the computational effort of which is small.

The discretized augmented Lagrangian now takes the form

$$L_c(q^{M_q, N}, u^{M_u, N}, \lambda^{M_u, N}) = \frac{1}{2}(\mathbf{u}^z)^T (S_\rho \otimes A_x^u) \mathbf{u}^z + \frac{\beta}{2}(\mathbf{q}^z)^T (S_\rho^{\text{mix}}) \mathbf{q}^z \\ + (\boldsymbol{\lambda}^z)^T S_{\rho, q} \mathbf{u}^z + \frac{c}{2}(\mathbf{e}^z)^T (S_\rho \otimes A_x^u) \mathbf{e}^z,$$

while the gradients (24) and (23) of  $L_c$  with respect to  $q$  and  $u$  are given by

$$D_q[L_c(q^{M_q, N}, u^{M_u, N}, \lambda^{M_u, N})] = \beta(S_\rho^{\text{mix}} \otimes A_x^q) \mathbf{q}^z + cS_{\rho, u} \mathbf{e}^z(\mathbf{q}, \mathbf{u}) \\ + S_{\rho, u} \boldsymbol{\lambda}^z - cS_{\rho, u} \mathbf{e}^z(\mathbf{f}) \quad (27)$$

and

$$D_u[L_c(q^{M_q, N}, u^{M_u, N}, \lambda^{M_u, N})] = (S_\rho \otimes A_x^u) \mathbf{u}^z + cS_{\rho, q} \mathbf{e}^z(\mathbf{q}, \mathbf{u}) \\ + S_{\rho, q} \boldsymbol{\lambda}^z - cS_{\rho, q} \mathbf{e}^z(\mathbf{f}) \quad (28)$$

respectively. The auxiliary problems ( $P_{\text{aux}}^q$ ) and ( $P_{\text{aux}}^u$ ) whose solutions yield updates for the parameter  $q$  as well as the state  $u$ , can therefore be discretized in the form of two linear systems of size  $M^q N$  and  $M^u N$  respectively. These systems are where the bulk of the computational effort is spent. In our numerical computations, we employ the preconditioned conjugate gradient method.

## 7 Numerical Results

In this section, we discuss three numerical examples to illustrate the use of the augmented Lagrangian method to estimate the statistical distribution of a spatially varying diffusion parameter  $q$  from the measured output  $\hat{u}$ . In each case, we compute sample paths of  $\hat{u}$  by solving (1) using sample paths of the exact parameter  $q$  and a deterministic forcing term  $f$ , and perturbing the result slightly to account for measurement variability. We use a hierarchical basis of piecewise linear hat functions of the same order  $L$  to interpolate  $q, u, \lambda$  and  $\hat{u}$ . For the first two examples, the random variables that define the uncertain parameter are also used to express the model output and we construct the stochastic interpolant of  $\hat{u}$  directly from that of  $q$  by generating its sample paths at the appropriate sparse grid nodes. For the third example, we first compute a truncated KL expansion of  $\hat{u}$ , based on a randomly generated sample, and estimate the joint density of the pertinent random variables from which we then compute an interpolant. Throughout, we use the augmented Lagrangian with parallel splitting to effect the minimization. For the sake of regularization, we use a spatial discretization of  $\hat{u}$  that is twice as fine as that of  $q$  throughout. To assess the accuracy of our approximation, we compare the first few central moments of  $q$  with those of its approximation  $\hat{q}$ . In these examples, we did not enforce positivity of the constraint explicitly.

**Example 7.1.** The first example serves to demonstrate the augmented Lagrangian method for a problem in 1 spatial- and 4 stochastic dimensions. The exact parameter  $q$  and deterministic forcing term  $f$  are defined over the domain  $[0, 1]$  by

$$q(x, y) = 2 + x^2 + \frac{1}{2} \sum_{i=1}^4 \cos(i\pi x) Y_i(\omega), \quad Y_i(\omega) \sim \text{i.i.d Uniform}([0, 1]), \quad \text{and} \\ f(x, y) = 6x^2 - 2x + 4$$

respectively. The manufactured solution  $\hat{u}$  is perturbed by uniform random noise of relative size  $\delta = 0.001$ . We use 30 elements for  $q$  and 60 for  $u, \hat{u}$ , and  $\lambda$ , a regularization term  $\beta =$

5e-5, an initial guess  $q_0 = 1$ , and terminate the program when the norm of the difference of successive iterates is within the tolerance 1e-5. Both sub-problems (27) and (28) are solved using a conjugate gradient routine with a relative residual tolerance of 1e-5. For this example, it is possible to plot and compare the sample paths of  $q$  and  $\hat{q}$  at the collocation points. Figure 1 shows that qualitatively, they indeed look similar. In Figure 2, we compare the first 4 central moments of  $q$  and  $\hat{q}$ , which confirms that we are able to identify the statistical behavior of  $q$  with a high accuracy (well within the magnitude of the noise added to the data). Table 1 summarizes the convergence behavior of the algorithm.

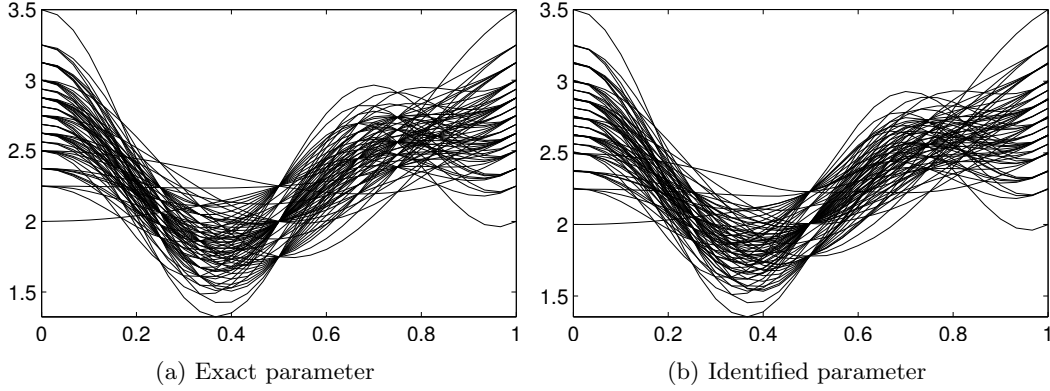


Figure 1: Sample paths of the exact- and identified parameter.

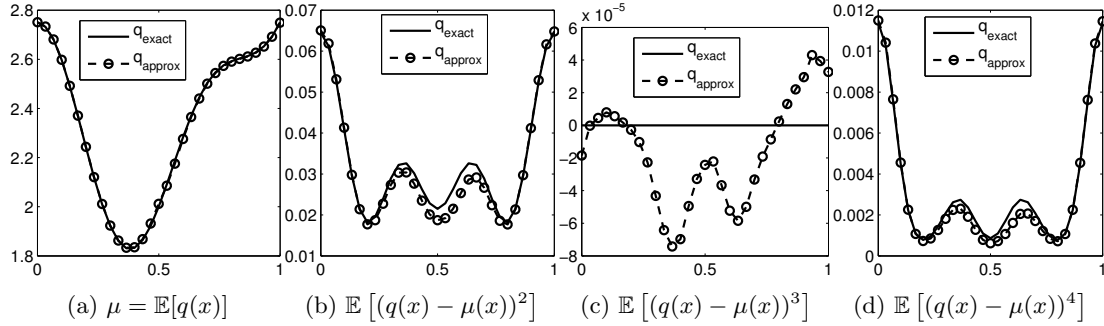


Figure 2: The first 4 central moments of  $q$  and of its approximation  $\hat{q}$ .

**Example 7.2.** As for deterministic inverse problems, the parameter  $q$  may not be identifiable in certain spatial regions, due to the shape of the output for instance (see [22]). This example investigates the role of regularization in this context. We chose a random output  $\hat{u}$ , most of whose sample paths have a zero gradient over a large area. Specifically, the deterministic forcing term  $f$  is given by

$$f(x_1, x_2) = -\nabla \cdot (k(x_1, x_2) \nabla (w(x_1)w(x_2))),$$

where

$$w(x) = \begin{cases} 9x^2 + 6x, & x \in [0, 1/3] \\ 1, & x \in (1/3, 2/3) \\ -9x^2 + 12x - 3, & x \in [2/3, 1] \end{cases},$$



Step	PCG Iterations		$L^2$ error	Increments	Cost Functional
	$(P_{\text{aux}}^q)$	$(P_{\text{aux}}^u)$	$\ q - \hat{q}\ _{L^2}$	$\ \hat{q}_k - \hat{q}_{k-1}\ _{L^2}$	$J(q_k, u_k, \lambda_k)$
1	1737	1246	1.9039	-	1.7764e-20
2	86	328	6.7864e-05	1.9019	5.4329e-05
3	25	118	9.2998e-05	2.7416e-06	5.3453e-05

Table 1: Computational work and convergence diagnostics for for Example 7.1.

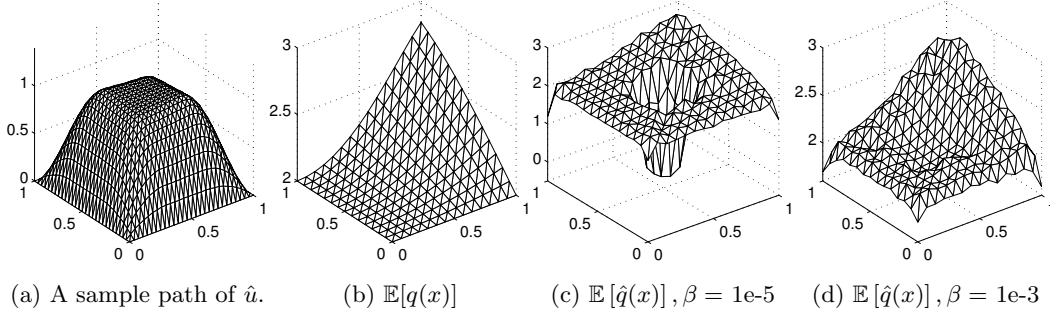
and

$$k(x_1, x_2) = 2 + \sin(x_1^2 x_2).$$

The exact parameter  $q$  is given by

$$q(x_1, x_2, Y_1, Y_2, Y_3) = 2 + \sin(x_1^2 x_2) + \frac{1}{8} \sum_{i=1}^3 \sin(i\pi x_1) \sin(i\pi x_2) Y_i,$$

where  $Y_i \sim \text{i.i.d. Uniform}([-1, 1])$ ,  $i = 1, 2, 3$ . We computed its approximation  $\hat{q}$  on a uniform triangular mesh of 392 elements over the unit square, added the same level of noise  $\delta$  as before, and interpolated in the stochastic component at level  $L = 4$ . Figure 3a shows a typical sample path of  $\hat{u}$ . The problem was first solved using a regularization parameter  $\beta = 1\text{e-}5$ , then again using  $\beta = 1\text{e-}3$ . In both cases the convergence tolerance was set to  $1\text{e-}4$  and the conjugate gradient tolerance was  $1\text{e-}5$ .



Figures 3b, 3c, and 3d show the mean of  $q$  and of  $\hat{q}$  in each of these cases. Using a larger regularization parameter penalizes steep gradients, thereby improving the conditioning of the inverse problem, albeit at the cost of accuracy. Evidently, regularization continues to play a significant role in the estimation of uncertain parameters. Similar figures can be plotted for the higher order moments. Quantitative outputs of the algorithm are provided in Table 2.

**Example 7.3.** For this example, the random variables used to express the identified parameter are estimated from sample paths of the model output  $\hat{u}$ . The deterministic forcing term satisfies

$$f(x_1, x_2) = -\nabla \cdot ((4 + x_1 x_2) \nabla \sin(\pi x_1) \sin(\pi x_2)),$$

while

$$\begin{aligned} q(x_1, x_2, y_1, y_2, y_3) = & 4 + x_1 x_2 + 0.5 \sin(\pi x_1) \sin(\pi x_2) Y_1 \\ & + 0.25 \cos(0.5\pi x_1) \sin(0.5\pi x_2) Y_2 + 0.25 \cos(\pi x_1) \cos(\pi x_2) Y_3, \end{aligned}$$

Step	$L^2$ error	Increments	Cost Functional	AL Functional
	$\ q - \hat{q}\ _{L^2}$	$\ \hat{q}_k - \hat{q}_{k-1}\ _{L^2}$	$J(q_k, u_k, \lambda_k)$	$L(q_k, u_k, \lambda_k)$
0	1.4083	-	-2.1871e-17	0.0463
1	0.0225	1.2559	0.0102	0.0130
2	0.0054	9.2e-3	0.0115	0.0118
3	0.0043	4.6058e-04	0.0117	0.0117
4	0.0043	5.2789e-05	0.0116	0.0116

Table 2: Convergence table for Algorithm 2 applied to Example 7.2 with  $\beta = 1e-3$ .

where  $Y_i \sim \text{i.i.d. Uniform}([-1, 1])$ . Using random samples of these input parameters, we generated 1000 sample paths of  $\hat{u}$ , which we then decomposed according to the method outlined in Section 6. No additional noise was added to the sample paths. For this problem, 2 KL expansion terms suffice to represent the sample  $\hat{u}$  so that the remaining expansion terms contribute less than  $\text{tol} = 1e-7$  to the field's variance. We express each random variable  $Y_i, i = 1, 2$  as the inverse image of a uniform random variable under its empirical cumulative distribution function (cdf). The appropriate graphs are shown in Figure 3.

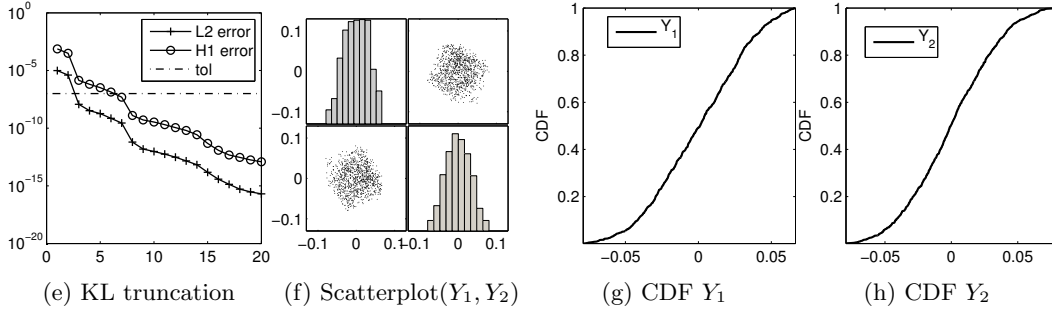


Figure 3: Sparse grid interpolation of  $\hat{u}$  based on a random sample of 1000 paths.

As in Example 7.2, we discretize  $q$  using a uniform spatial mesh of 392 elements. In addition, we choose a sparse grid interpolation level  $L = 4$ . We use a regularization term  $\beta = 1e-5$ , and terminate the optimization algorithm when the  $L^2$  norm of successive iterates is within the tolerance level of  $1e-5$ . For the conjugate gradient subroutines, we use a tolerance of  $1e-6$ . As before, we compare the central moments of the identified parameter  $\hat{q}$  with those of its exact counterpart  $q$  to assess its accuracy. Figure 4 shows that, qualitatively, the estimate is good. Since the random variables used to express  $\hat{q}$  differ from  $Y_1$  and  $Y_2$ , it is impossible to compute the exact error as part of the optimization run. We nevertheless record relevant convergence diagnostics in Table 3.

## 8 Conclusion

In this paper we have formulated a fairly general variational framework for the estimation of spatially distributed, uncertain diffusion coefficients in stationary elliptic problems, based on statistical measurements of the model output. In contrast to the Bayesian approach, we used a parametrization of the coefficient in terms of a finite number of variables, allowing us to not

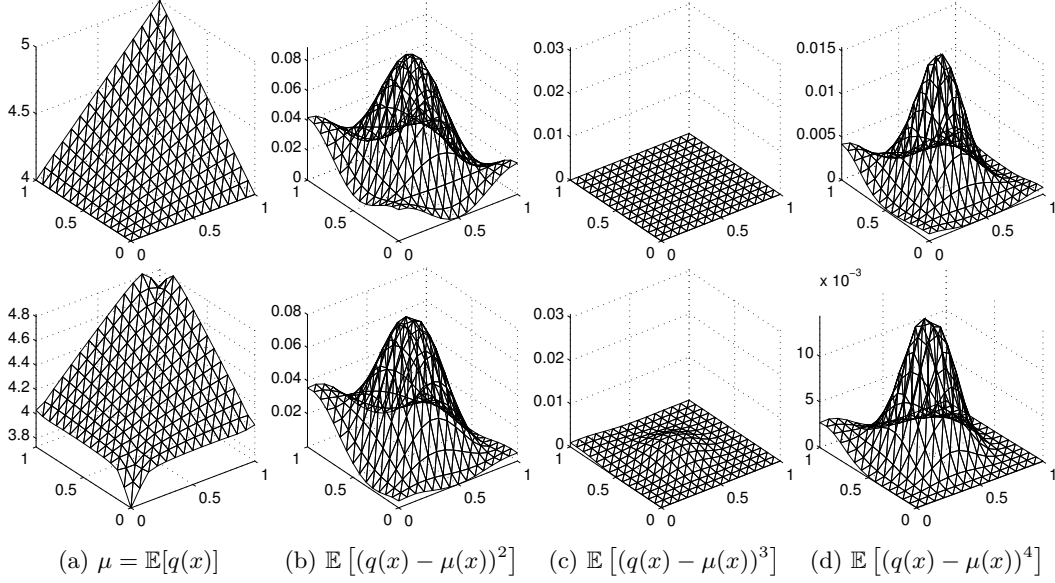


Figure 4: The first 4 central moments of the exact parameter  $q$  (top row) and those of the identified parameter  $\hat{q}$  (bottom row).

Step	Increments	Cost Functional	AL Functional
	$\ \hat{q}_k - \hat{q}_{k-1}\ _{L^2}$	$J(q_k, u_k, \lambda_k)$	$L(q_k, u_k, \lambda_k)$
0	-	-	-
1	10.6826	6.1409e-05	6.1461e-05
2	3.7636e-05	5.3654e-05	5.3680e-05
3	1.2276e-05	5.0058e-05	5.0075e-05
4	4.3860e-06	4.8018e-05	4.8029e-05

Table 3: Convergence table for Algorithm 2 applied to Example 7.3.

only estimate the statistical mismatch between the predicted- and observed output, but also to determine the perturbations of  $q$  that will result in a decrease in the degree of mismatch. In light of the potential size in the number of degrees of freedom, the computation of quantities such as steepest descent directions, or cost functional evaluations may require considerable computational cost. We are currently investigating ways to reduce the computational overhead, through parallelization [40], multigrid methods, or the use of sensitivity information [9].

## References

- [1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, vol. 140 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, second ed., 2003.
- [2] I. BABUŠKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 1005–1034.

- [3] I. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 800–825.
- [4] ———, *Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation*, Computer Methods in Applied Mechanics and Engineering, 194 (2005), pp. 1251–1294.
- [5] H. T. BANKS AND K. L. BIHARI, *Modelling and estimating uncertainty in parameter estimation*, Inverse Problems, 17 (2001), pp. 95–111.
- [6] H. T. BANKS AND K. KUNISCH, *Estimation Techniques for Distributed Parameter Systems*, vol. 1 of Systems & Control: Foundations & Applications, Birkhäuser Boston Inc., Boston, MA, 1989.
- [7] V. BARTHELMANN, E. NOVAK, AND K. RITTER, *High dimensional polynomial interpolation on sparse grids*, Advances in Computational Mathematics, 12 (2000), pp. 273–288.
- [8] A. BINDER, H. W. ENGL, C. W. GROETSCH, A. NEUBAUER, AND O. SCHERZER, *Weakly closed nonlinear operators and parameter identification in parabolic equations by tikhonov regularization*, Applicable Analysis, 55 (1994), pp. 215–234.
- [9] J. BORGGAARD, V. L. NUNES, AND H.-W. VAN WYK, *Sensitivity and uncertainty quantification of random distributed parameter systems*, Mathematics in Engineering, Science and Aerospace, 4 (2013), pp. 117–129.
- [10] H.-J. BUNGARTZ AND S. DIRNSTORFER, *Multivariate quadrature on adaptive sparse grids*, Computing, 71 (2003), pp. 89–114.
- [11] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 147–269.
- [12] T. F. CHAN AND X.-C. TAI, *Identification of discontinuous coefficients in elliptic problems using total variation regularization*, SIAM Journal on Scientific Computing, 25 (2003), pp. 881–904 (electronic).
- [13] ———, *Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients*, Journal of Computational Physics, 193 (2004), pp. 40–66.
- [14] Z. CHEN AND J. ZOU, *An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems*, SIAM Journal on Control and Optimization, 37 (1999), pp. 892–910.
- [15] T. GERSTNER AND M. GRIEBEL, *Numerical integration using sparse grids*, Numerical Algorithms, 18 (1998), pp. 209–232.
- [16] M. R. HESTENES, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4 (1969), pp. 303–320.
- [17] K. ITO, M. KROLLER, AND K. KUNISCH, *A numerical study of an augmented Lagrangian method for the estimation of parameters in elliptic systems*, SIAM Journal on Scientific and Statistical Computing, 12 (1991), pp. 884–910.
- [18] K. ITO AND K. KUNISCH, *The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces*, Mathematical Programming, 46 (1990), pp. 341–360.
- [19] K. ITO AND K. KUNISCH, *The augmented Lagrangian method for parameter estimation in elliptic systems*, SIAM Journal on Control and Optimization, 28 (1990), pp. 113–136.
- [20] J. KLEMELÄ, *Smoothing of Multivariate Data: Density Estimation and Visualization*, vol. 737, John Wiley & Sons, 2009.
- [21] K. KUNISCH AND X.-C. TAI, *Sequential and parallel splitting methods for bilinear control problems in hilbert spaces*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 91–118.

- [22] K. KUNISCH AND L. W. WHITE, *Identifiability under approximation for an elliptic boundary value problem*, SIAM Journal on Control and Optimization, 25 (1987), pp. 279–297.
- [23] M. LOÈVE, *Probability Theory. II*, Springer-Verlag, New York, fourth ed., 1978. Graduate Texts in Mathematics, Vol. 46.
- [24] X. MA AND N. ZABARAS, *An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations*, Journal of Computational Physics, 228 (2009), pp. 3084–3113.
- [25] ———, *An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations*, Journal of Computational Physics, 229 (2010), pp. 3884–3915.
- [26] H. MAURER AND J. ZOWE, *First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems*, Mathematical Programming, 16 (1979), pp. 98–110.
- [27] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2309–2345.
- [28] E. NOVAK AND K. RITTER, *High dimensional integration of smooth functions over cubes*, Numerische Mathematik, 75 (1996), pp. 79–97.
- [29] M. J. D. POWELL, *A fast algorithm for nonlinearly constrained optimization calculations*, in Numerical analysis (Proc. 7th Biennial Conf., Univ. Dundee, Dundee, 1977), Springer, Berlin, 1978, pp. 144–157. Lecture Notes in Math., Vol. 630.
- [30] M. REED AND B. SIMON, *Methods of Modern Mathematical Physics. I*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second ed., 1980. Functional analysis.
- [31] R. T. ROCKAFELLAR, *Coherent approaches to risk in optimization under uncertainty*, in In Tutorials in Operations Research INFORMS, 2007, pp. 38–61.
- [32] A. SANDU, *Solution of Inverse Problems using Discrete ODE Adjoint*, John Wiley and Sons, Ltd, 2010, pp. 345–365.
- [33] C. SCHWAB AND R. A. TODOR, *Karhunen-Loève approximation of random fields by generalized fast multipole methods*, Journal of Computational Physics, 217 (2006), pp. 100–122.
- [34] D. W. SCOTT, *Multivariate Density Estimation*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1992. Theory, practice, and visualization, A Wiley-Interscience Publication.
- [35] D. W. SCOTT AND S. R. SAIN, *Multi-dimensional density estimation*, Handbook of Statistics, 24 (2005), pp. 229–261.
- [36] S. A. SMOLYAK, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, in Dokl. Akad. Nauk SSSR, vol. 4, 1963, p. 123.
- [37] A. M. STUART, *Inverse problems: a Bayesian perspective*, Acta Numerica, 19 (2010), pp. 451–559.
- [38] A. TARANTOLA, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [39] V. N. TEMLYAKOV, *On approximate recovery of functions with bounded mixed derivative*, Journal of Complexity, 9 (1993), pp. 41–59.
- [40] H.-W. VAN WYK, *Identification of uncertain, spatially varying parameters through multilevel sampling*, in Proceedings of the 19th IFAC World Congress, 2014.

- [41] D. XIU AND G. E. KARNIADAKIS, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM Journal on Scientific Computing, 24 (2002), pp. 619–644 (electronic).
- [42] N. ZABARAS AND B. GANAPATHYSUBRAMANIAN, *A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach*, Journal of Computational Physics, 227 (2008), pp. 4697–4735.